State-dependent attention and pricing decisions*

Javier Turen[†]

February 2022

Abstract

This paper studies price-setting decisions under Rational Inattention. Prices are set by tracking an unobserved target whose distribution is also *unknown*. Information acquisition is dynamic and fully flexible since, given information acquired previously, price-setters choose the amount of information they collect as well as how they want to learn about both the outcome and its distribution. We show that by allowing for imperfect information to be the unique source of rigidity, the model can reconcile stylized facts in the microeconomic evidence on price setting while *simultaneously* being consistent with empirical results on state-dependent attention.

KEYWORDS: *Rational Inattention, Dynamic Information, Price Dispersion, Endogenous Attention.*

JEL CLASSIFICATION: E31, E32, D82, D83.

*I am extremely grateful to Raffaella Giacomini and Vasiliki Skreta for invaluable guidance and advices. I am also grateful for the useful suggestions and comments of Filip Matejka, Antonio Cabrales, Morten Ravn, Mirko Wiederholt, Franck Portier, Ralph Luetticke, Nicolas Figueroa, Edouard Schaal, Carlos Carvalho, Roger E.A. Farmer, Ernesto Pasten, Juan Pablo Torres-Martinez, Nezih Guner, Davide Melcangi, Pau Roldan, Alejandro Vicondoa, Silvia Sarpietro and Carlo Galli. An earlier version of this paper was entitled "Rational Inattention-driven dispersion with volatility shocks". I am also grateful to Fondecyt-Anid for funding (project *Fondecyt Iniciación* No 11190753). All errors and omissions are mine.

†Instituto de Economía, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, 7820436, Santiago, Chile, (email: jturen@uc.cl)

1 Introduction

Firms are regularly making decisions under dynamically evolving circumstances. Most of these decisions are made with only partial information about unpredictable components of their demand, industries, or the economy. Recent evidence has supported this view by stressing that firm managers' beliefs are consistent with theories of information rigidity, particularly with models of inattention, Coibion, Gorodnichenko and Kumar (2018). The evidence suggests that the degree of this rigidity is state-dependent, which is consistent with agents *deciding* to acquire more information in response to a less predictable economic environment, Coibion and Gorodnichenko (2015). While intuitive, there still scarce evidence on the implications of time-varying imperfect information for firm decisions. In particular, little is known about the extent to which models of information frictions can address the empirical evidence documented across the business cycle.

This paper studies price-setting decisions under dynamic imperfect information. We argue that the information friction alone can explain several features of the micro price data while *simultaneously* being consistent with the empirical evidence of state-dependent attention. This is the main contribution of the paper. While the stylized facts of micro price data have been successfully matched using state-dependent pricing models, Klenow and Kryvtsov (2008), Nakamura and Steinsson (2008) and Vavra (2013), this paper sheds new light by studying further attention-related moments, such as the ones estimated by Coibion and Gorodnichenko (2015) (henceforth, CG (2015)), that were not incorporated in previous work.

We propose a model able to replicate several stylized facts of price changes, at both the cross-sectional and time series level, while being simultaneously consistent with the evidence regarding the information acquisition process of agents over the business cycle. By construction a stylized price-rigidity model, such as a menu-cost model, is muted when it comes to generating relevant features about agents' time-varying beliefs. This paper highlights the implications of information frictions on aggregate conditions by mapping the role of dynamic imperfect information on aggregate price stability. The results support the importance of this friction as a relevant constraint faced by firms. Moreover, they also contribute to the discussion about optimal policy within a price-setting framework.[1]

The model features endogenous attention with costly entropy reduction and is used to study how firms set prices when the distribution of shocks is time-varying. The model follows the literature on "Rational Inattention" (henceforth, RI) Sims (2003), and allows for a dynamic and fully flexible information scheme. While past acquired information is relevant, we do not impose further assumptions on the *amount* of information being acquired or how owners *choose* to acquire it, i.e., there are no parametric assumptions on the distribution of signals. Firms collect information to update their beliefs about the realization of an aggregate fundamental,

[1]In particular, Paciello and Wiederholt (2013) shows how replacing price rigidity with an imperfect information mechanism is not innocuous for policy counterfactuals.

along with the distribution that generated it. This is relevant as the distribution of shocks is unknown and likely to change over time, reflecting unanticipated periods of lower or higher uncertainty such as recessions. As the predictability of the outcome depends on the persistent parameters that govern the distribution, the incentives to acquire information respond to price-setters idiosyncratic beliefs, creating a dynamic learning problem.

Regarding the empirical evidence, the model generates a positive response of price-change dispersion to volatility shocks along with a positive comovement between the dispersion and frequency of price changes.[2] Bachmann, Born, Elstner and Grimme (2019), Drenik and Perez (2018) and Klepacz (2017) document the existence of a positive correlation between volatility shocks and price-change dispersion. The positive correlation between price-change dispersion (intensive margin) and the frequency of price changes (extensive margin) was shown by Vavra (2013). As owners are active learners, the results are also consistent with the aforementioned presence of time-varying attention. In particular, the model replicates the dynamic patterns of information acquisition over the business cycle consistent with the results presented by CG (2015).

The model's ability to replicate the aforementioned evidence depends on the simultaneous presence of a dynamic, flexible information framework, with time-invariant heterogeneous costs. The model then nests two previously studied settings in the RI literature. With full information about the shock distribution, the model turns into a more stylized static setup with a Gaussian unobserved target-price. This setting resembles the one presented by Woodford (2003) and Maćkowiak and Wiederholt (2009). A dynamic model with homogeneous information costs was also analyzed by Matějka (2015). We show that each assumption on its own is not enough to fully replicate the dynamic relationships suggested by the data for prices and information acquisition.

While imperfect information is the only friction, the model is still consistent with the fact that prices stay constant for a certain amount of time. As argued by Matějka (2015) and Jung, Kim, Matějka and Sims (2019), a rationally inattentive agent chooses to price discretely when the processes for the unobserved shocks are not Gaussian. Uncertainty about the correct distribution is modeled by assuming a mixture of normal distributions for the optimal price. Because of this assumption, agents do not change their prices every period, creating price stickiness. The simulated duration is, however, *shorter* compared to the data. Adding further frictions, such as price rigidities within the described dynamic learning structure, emerges as a natural extension of this paper.

Solving a model in which information acquisition is dynamic and fully flexible imposes several methodological challenges. Acquired information has an effect on both pricing decisions and posterior beliefs about the next period's distribution. To allow for a dynamic setting,

[2] We make a distinction between "dispersion" and "volatility." In this context, dispersion refers to the spread (typically measured as the standard deviation) of endogenous variables for firms' cross-section. Meanwhile, volatility refers to the spread of exogenous shocks.

a common assumption in the RI literature is to assume a Gaussian distribution for the shock process, which is known with certainty.[3] In this paper, however, the optimal signal structure depends on how firms choose to attach different probabilities to each possible shock's distribution. As information acquired in the past guides their decisions, the challenge is how to characterize the effects of current flexible information on posterior beliefs. We circumvent this problem by building on the solution proposed by Steiner, Stewart and Matějka (2017), henceforth SSM (2017). Furthermore, we provide an algorithm tailored to solve this dynamic learning model.

Literature Review. The paper contributes to the price-setting literature with information frictions. Alvarez, Lippi and Paciello (2011) solves a price-setting problem with observation and menu costs. The authors show how these two costs complement each other, delivering different implications for the timing of price reviews. Moscarini (2004) introduces a pricing problem with limited information, where agents are restricted to infrequently receiving new information, creating inertia in their behavior. Woodford (2009) introduces a setting with menu-costs, where the decisions to conduct a price review is made under RI. Vavra (2013) studies the dynamic behavior of price setting and argues how a menu-cost model, with time-variant idiosyncratic shocks, can match the dynamic features of prices. While all these papers rely on price rigidities' crucial role, this paper aims to highlight the role of information rigidities as a key driver behind aggregate decisions. Gorodnichenko (2008) solves a model with information frictions and menu costs, which also predicts that total acquired information endogenously increases after a surge in aggregate uncertainty. Our results extend this finding by showing that a price-setting model with the characteristics mentioned above is also consistent with micro price data. Baley and Blanco (2018) studies dynamic pricing with menu-costs and information rigidities. In their set up, the timing of volatility shocks is known with certainty, which is precisely the primary assumption this paper aims to relax.

RI models have proven useful when rationalizing the empirical behavior of micro prices along with their aggregate implications. Maćkowiak and Wiederholt (2009) propose a pricing model with endogenous attention to explain the sluggish response of prices to aggregate shocks. Matějka (2015) alternatively introduce a model that does not rely on quadratic objectives nor Gaussian distributions, as in Maćkowiak and Wiederholt (2009), which endogenously generates price discreteness. Afrouzi (2018) solves a dynamic general equilibrium model with inattentive price setters, Gaussian signals, and strategic complementarities between them. Paciello and Wiederholt (2013) shows how under costly information, monetary policy can reduce inefficient price dispersion by affecting the response of profit-maximizing prices to unobserved markup shocks. Finally, Stevens (2019) presents a price-setting model with constrained information, which captures the heterogeneous patterns of adjustments observed in the data, along with the sluggish response of prices to shocks. This paper contributes to this literature by studying the

[3] This assumption combined with a quadratic loss function leads to a closed-form for the optimal signal structure, given by the outcome realization plus normally distributed noise as in Woodford (2003) and Maćkowiak, Matějka and Wiederholt (2018).

unexplored ability of these models to match the aggregate implications of the two price margins while being consistent with the evidence at the micro-level.

The rest of the paper is structured as follows. In Section 2, we introduce the model set up and discuss the dynamic costly information setting. We then fully derive and characterize the solution to the problem. Section 3 presents the algorithm used to replicate both cross-sectional and time-series moments from the data. The main results of the paper are discussed in Section 4, where we lay out both the individual and aggregate implications under persistent volatility shocks. Section 5 introduces some alternative specification for the model. Finally, Section 6 concludes.

2 The dynamic learning pricing model

2.1 Set up

The setting is a partial equilibrium model where time is discrete $t \geq 0$ and there are a fixed number of firms $i = 1, \ldots, N$. Firm owners choose prices p_{it} from a finite set Ω_p to maximize the present discounted value of profits. Each firm can adjust its price costlessly in every period so p_{it} is set to maximize current profits $\widehat{\Pi}(p_{it}, \widehat{p}_{it})$. Following Caplin and Leahy (1997) and Alvarez et al. (2011), the profit function is set equal to:

$$\widehat{\Pi}(p_{it}, \widehat{p}_{it}) = \gamma(p_{it} - \widehat{p}_{it})^2 \tag{1}$$

The objective function (1) can be interpreted as a second-order approximation of a more general profit function around its non-stochastic steady state. The details behind the approximation are presented in Appendix 7.1. The parameter γ represents the curvature of the demand function and $\widehat{p}_{it}$ is labelled as the idiosyncratic "price-target". Given the approximation, $\widehat{p}_{it}$ is a function of firms' marginal costs. In the model, owners do not have complete information about cost conditions as they cannot fully track the shocks affecting their production due to their own limitations in processing information.[4]

Imperfect information about the current distribution of $\widehat{p}_{it}$ is modeled in the following way. There are two independent shocks drawn in each period t from finite sets, $\sigma_t \in \Omega_\sigma$ and $\epsilon_{it} \in \Omega_\epsilon$. The price-target is assumed equal to $\widehat{p}_{it} = \sigma_t \epsilon_{it}$. Underlying the shocks evolution is a probability distribution induced by a Markov Chain on Ω_σ and a discretized Gaussian on Ω_ϵ, with mean zero

[4] Bachmann and Moscarini (2011) argues how different cost variables (such as input price elasticities or costs structures) are hard to estimate by firms. Think about owners who want to maximize profits but have multiple demands on their time such as reading reports about the firm's inventory levels, projecting future sales, testing and developing new products, collecting information about clients' reactions to historical prices, among others. Information is imperfect in this case, as owners cannot possibly precisely remember all the information when setting a price.

and unit variance. Thus, while the former shock is persistent, the latter is i.i.d. The stochastic process of both shocks is common information across firms. We assume $\Omega_\sigma := \{\sigma_L, \sigma_H\} \subseteq \mathbb{R}_+$, with $\sigma_H = \phi\sigma_L$, $\phi > 1$. The transition probabilities of switching from the σ_L to the σ_H state, and viceversa, are labelled as τ_{LH} and τ_{HL} respectively. As noticed, the unconditional process for the price-target is assumed stationary. Following SSM(2017), stationarity is needed to characterize the solution of a dynamic RI problem through a finite system of equations, as shown in Proposition 2 below.[5]

The two components of the target-price are aimed at capturing idiosyncratic and aggregate uncertainty across firms. Since neither of the two shocks is fully observed, firms are not only uncertain about the realization of $\widehat{p}_{it}$, they also do not know if the price was drawn from $N(0, \sigma_L^2)$ or $N(0, \sigma_H^2)$. Therefore we will now refer interchangeably to persistent volatility states and different distributions for the target-price throughout the paper.

2.2 Information Acquisition

To optimally set prices firms collect information about $\widehat{p}_{it}$ to minimize (1). As the predictability of $\widehat{p}_{it}$ is "state-dependent", owner's beliefs about its current distribution will discipline their efforts to collect information. The learning process is dynamic in the sense that previously acquired information is still relevant for current decisions due to the persistence of volatility states.

Owners acquire information about $\widehat{p}_{it}$ by choosing a signal $s_{it} \in \Omega_s$, where $|\Omega_p| \leq |\Omega_s|$. Firms are rationally inattentive since through costly information they aim to reduce the entropy of their beliefs, Sims (2003). Entropy about $\widehat{p}_{it}$ is defined as $\mathcal{H}(\widehat{p}_{it}|s_i^{t-1}) \equiv E[-log(\widehat{p}_{it})|s_i^{t-1}]$, where $s_i^{t-1} \equiv \{s_{it-1}, s_{it-2}, \ldots, s_{i0}\}$. Then s_i^{t-1} is the information set generated by the history of signals from firm i up to $t-1$. Prior uncertainty about $\widehat{p}_{it}$ is then $\mathcal{H}(\widehat{p}_{it}|s_i^{t-1}) = -\sum_{\sigma_t}\sum_{\epsilon_{it}} g_{it}(\widehat{p}_{it}|s_i^{t-1}) log(g_{it}(\widehat{p}_{it}|s_i^{t-1}))$.

In line with RI models, the reduction in uncertainty is quantified by Shannon (1948)'s measure of mutual information flow:

$$\mathcal{I}(\widehat{p}_{it}, s_{it}|s_i^{t-1}) \equiv \mathcal{H}(\widehat{p}_{it}|s_i^{t-1}) - E_{s_{it}}[\mathcal{H}(\widehat{p}_{it}|s_{it})|s_i^{t-1}] \tag{2}$$

Information flow (2) is the difference between prior and posterior uncertainty about $\widehat{p}_{it}$, conditioned on lagged information s_i^{t-1}.[6]

[5] Based on the second order approximation, the target-price is equivalent to $log(P_{it}^*)$, where P_{it}^* is a constant mark-up over time-varying idiosyncratic marginal costs C_t, see Appendix 7.1 for details. Besides the relevance of stationarity for the solution, our setup still allows for a persistent process for the marginal cost if we assume that $log(P_{it}^*) = \mu + t + \sigma_t\epsilon_t$ where the drift μ and the trend t are assumed common knowledge across firms.

[6] As described, the entropy formula relies on logarithms which depending on the base, changes the units by which we measure information. If the log is base two, then the information is measured in bits, while if it is e it is measured in nats.

Firms enter each period with prior beliefs $g_{it}(\widehat{p}_{it}|s_i^{t-1}) = m_{it}(\sigma_t|s_i^{t-1})h(\epsilon) \in \Delta(\Omega_{\widehat{p}})$ where $\Omega_{\widehat{p}} := \Omega_\sigma \times \Omega_\epsilon$. Hence, $\Delta(\Omega_{\widehat{p}})$ is the set of all probability distributions on $\Omega_{\widehat{p}}$. In the definition, $m_{it}(\sigma_t|s_i^{t-1})$ and $h(\epsilon)$ are the prior probability measures of σ_t and ϵ_{it} respectively. Since the probability of $\epsilon_{it} \in \Omega_\epsilon$ is i.i.d. and its stochastic process is known, its prior probability is constant across firms and independent of the history of signals.

During each period t, owners choose an "information strategy" $f_{it}(s_{it}, \widehat{p}_{it}|s_i^{t-1}) \in \Delta_{g_{it}}(\Omega_s \times \Omega_{\widehat{p}})$ and a "pricing strategy" $p_{it} : \Delta(\Omega_{\widehat{p}}) \to \Omega_p$. Information acquisition is then summarized by the joint probability distribution of signals and optimal prices where $\Delta_{g_{it}}(\Omega_s \times \Omega_{\widehat{p}})$ is the set of all probability distributions on $\Omega_s \times \Omega_{\widehat{p}}$, consistent with prior beliefs $g_{it}(\widehat{p}_{it}|s_i^{t-1})$. After the price-target is drawn, the choice of $f_{it}(s_{it}, \widehat{p}_{it}|s_i^{t-1})$ reflects the type of acquired signal based on the information simplification process chosen by each owner and given information acquired in the past.

The following proposition shows that the expression for $\mathcal{I}(\widehat{p}_{it}, s_{it}|s_i^{t-1})$ in (2) can be written as a function of $f_{it}(s_{it}, \widehat{p}_{it}|s_i^{t-1})$.

Proposition 1 : ***Mutual Information Equivalence***

Shannon's mutual information (2) is equal to:

$$\mathcal{I}(\widehat{p}_{it}, s_{it}|s_i^{t-1}) = \sum_{s_{it}} \sum_{\sigma_t} \sum_{\epsilon_{it}} f_{it}(s_{it}, \widehat{p}_{it}|s_i^{t-1}) log \left(\frac{f_{it}(s_{it}, \widehat{p}_{it}|s_i^{t-1})}{g_{it}(\widehat{p}_{it}|s_i^{t-1}) f_{it}(s_{it}|s_i^{t-1})} \right) \tag{3}$$

Proof in Appendix 7.2.

By setting an information strategy $f_{it}(s_{it}, \widehat{p}_{it}|s_i^{t-1})$, owners are choosing the total amount of information $\mathcal{I}(\widehat{p}_{it}, s_{it}|s_i^{t-1})$ to acquire during each period.

2.3 The problem in two stages

Let us discuss the timing of the model. Within each period owners face two decisions: given prior beliefs $g_{it}(\widehat{p}_{it}|s_i^{t-1})$ they choose $f_{it}(s_{it}, \widehat{p}_{it}|s_i^{t-1})$ and then, endowed with this new information, they set prices p_{it}^*. Owners are Bayesian as by combining posterior beliefs about σ_t with τ_{LH} and τ_{HL}, they form prior beliefs for the next period $g_{it+1}(\widehat{p}_{it+1}|s_{it}, s_i^{t-1}) = m_{it+1}(\sigma_{t+1}|s_{it}, s_i^{t-1})h(\epsilon)$.

The pricing strategy describes how owners react to the received signal s_{it} by mapping posterior beliefs $f_{it}(\widehat{p}_{it}|s_{it}, s_i^{t-1}) \in \Delta(\Omega_{\widehat{p}})$ to optimal prices $p_{it}^*(s_{it}|s_i^{t-1})$.

$$p_{it}^*(s_{it}|s_i^{t-1}) = \arg\max_{p_{it}} \sum_{\sigma_t} \sum_{\epsilon_{it}} \widehat{\Pi}(p_{it}, \widehat{p}_{it}) f_{it}(\widehat{p}_{it}|s_{it}, s_i^{t-1}) \tag{4}$$

At the information acquisition stage, owners face a trade-off. Signals with higher precision allow them to observe $\widehat{p}_{it}$ with less noise, where the precision is determined by the channel's capacity (3). While owners can constantly modify the capacity, the cost of each additional unit of information is given by $\lambda_i > 0$. The cost directly affects the profit function, and it is assumed to differ across firms.

Information is fully flexible as firms set the precision of their signals by choosing $f_{it}(\widehat{p}_{it}, s_{it}|s_i^{t-1})$ without adding further parametric assumption about its particular shape. The chosen form for the joint probability distribution determines total acquired information for each moment of time. Moreover, as states are unobserved, each information strategy ultimately depend on owner's *perceived* prior distribution for $\widehat{p}_{it}$:

$$\widehat{p}_{it} \sim m_{it}(\sigma_L|s_i^{t-1})N(0, \sigma_L^2) + (1 - m_{it}(\sigma_L|s_i^{t-1}))N(0, \sigma_H^2)$$

Where $m_{it}(\sigma_L|s_i^{t-1})$ is the prior probability attached to the low volatility state of firm i, given information acquired in the past.

At the first stage, given the policy function $p^*_{it}(s_{it}|s_i^{t-1})$, $\mathcal{I}(\widehat{p}_{it}, s_{it}|s_i^{t-1}) \geq 0$ and $g_{it}(\widehat{p}_{it}|s_i^{t-1})$, firms choose the conditional distribution $f_{it}(s_{it}|\widehat{p}_{it}, s_i^{t-1})$ to maximize expected profits relative to the cost of information:

$$f_{it}(s_{it}|\widehat{p}_{it}, s_i^{t-1}) = \arg \max_{\widehat{f}(.) \in \Delta_g(\Omega_s)} \sum_{s_{it}} \sum_{\sigma_t} \sum_{\epsilon_{it}} \widehat{\Pi}(p^*_{it}, \widehat{p}_{it}) \widehat{f}_{it}(s_{it}|\widehat{p}_{it}, s_i^{t-1}) g_{it}(\widehat{p}_{it}|s_i^{t-1}) - \lambda_i \mathcal{I}(\widehat{p}_{it}, s_{it}|s_i^{t-1})$$

The information strategy shapes the posterior distribution of signals, which is equivalent to choosing $f_{it}(\widehat{p}_{it}, s_{it}|s_i^{t-1})$.

As the only purpose of costly information is to inform pricing decisions, through signals the firm is *implicitly* choosing its optimal price by determining $f(\widehat{p}_{it}|s_{it}, s_i^{t-1})$, i.e., we do not need to solve the endogenous choice of information separate from the pricing decisions of firms. Matejka and McKay (2014) and Matějka (2015) formally show this result for static RI problems, while SSM (2017) prove it within a dynamic setting with flexible information. Intuitively, each signal $s_{it} \in \Omega_s$ will be associated with just one price $p_{it} \in \Omega_p$. If two signals lead to the same price, and since entropy is a concave function, the firm could ended up setting the same price with a lower information cost.[7] Therefore it is enough to solve for the optimal distribution of prices conditioning on: (1) the realization of the target-price and (2) the history of previous pricing decisions $p_i^{t-1} := \{p_{it-1}, p_{it-2}, \dots p_{i0}\}$ instead of s_i^{t-1}.

[7] Moreover, since information is costly and $f(\widehat{p}_{it}, s_{it}|s_{it-1})$ is endogenous, necessarily $\mathcal{I}(\widehat{p}_{it}, p^*_{it}|s_{it-1}) \leq \mathcal{I}(\widehat{p}_{it}, s_{it}|s_{it-1})$. The linearity of the cost function is relevant under a dynamic setting as it prevents the firm from stocking unused information for future periods, SSM (2017).

2.3.1 A Markovian Solution

An optimal solution for the RI problem is defined as Markovian if the implied probability distribution that solves the dynamic problem at any time t depends on the last action p_{it-1} and the current state σ_t, but not on any earlier actions or states. Given the Markov switching process for σ_t and as shown by SSM(2017) and Miao and Xing (2019) (see Proposition 5 and 6, Section 4.2), a Markovian structure holds if the set of actions and states is finite, the Markov switching probabilities are time-invariant and if the proposed solution is interior. This latter condition means that all pricing decisions are chosen with positive probability at every point in time.

While assuming that the solution is interior, all the remaining conditions for a Markovian structure are satisfied in our model. Therefore, the relevant historical information is summarized by the lagged action p_{it-1} which, combined with the current optimal price $\widehat{p}_{it}$ allows us to characterize the dynamic learning model. Given the solution for the model, we will verify in Section 3.3 that it is indeed interior.

2.4 The dynamic RI problem

Let us now formally introduce the dynamic information acquisition problem. While the price-setting decision is static, the unobserved and persistent distribution of $\widehat{p}_{it}$ implies a correlation between consecutive periods. As the precision by which owners try to uncover the underlying state of the economy is subject to their choice, prior beliefs $m_{it}(\sigma_{jt}|p_{it-1}), j = L, H$ become the state variable of the problem.

During each period t, given $g_{it}(\widehat{p}_{it}|p_{it-1}) \in \Delta(\Omega_{\widehat{p}})$ and information costs $\lambda_i > 0$, owners choose $f_{it}(p_{it}, \widehat{p}_{it}|p_{it-1}) \in \Delta_g(\Omega_p \times \Omega_{\widehat{p}})$ to solve the dynamic problem:

$$V(m_{it}(\sigma_L|p_{it-1})) = \max_{f_{it}(p_{it},\widehat{p}_{it}|p_{it-1})} \sum_{\sigma_t} \sum_{\epsilon_{it}} \sum_{p_{it}} [\widehat{\Pi}(p_{it}, \widehat{p}_{it}) + \beta V(m_{it+1}(\sigma_L|p_{it}))] f_{it}(p_{it}, \widehat{p}_{it}|p_{it-1})$$

$$- \lambda_i \mathcal{I}(\widehat{p}_{it}, s_{it}|s_{it-1}) \tag{5}$$

Subject to:

$$\mathcal{I}(\widehat{p}_{it}, s_{it}|s_{it-1}) = f_{it}(p_{it}, \widehat{p}_{it}|p_{it-1}) log \left(\frac{f_{it}(p_{it}, \widehat{p}_{it}|p_{it-1})}{g_{it}(\widehat{p}_{it}|p_{it-1}) f_{it}(p_{it}|p_{it-1})} \right) \tag{6}$$

$$g_{it}(\widehat{p}_{it}|p_{it-1}) = m_{it}(\sigma_t|p_{it-1})h(\epsilon) = \sum_{p_{it}} f_{it}(p_{it}, \widehat{p}_{it}|p_{it-1}) \tag{7}$$

$$m_{it+1}(\sigma_L|p_{it}) \quad = \quad \mathcal{T}_{t+1}(f_{it}(\sigma_L|p_{it})) \tag{8}$$

$$0 \quad \leq \quad f_{it}(p_{it}, \widehat{p}_{it}|p_{it-1}) \leq 1 \tag{9}$$

Owners maximize the expected value of $\widehat{\Pi}(p_t, \widehat{p}_{it})$ with respect to the perceived probability distribution of p_{it} and $\widehat{p}_{it}$ relative to the total information cost. The cost λ_i forces the firm to form a probabilistic conjecture of its optimal price given both the unobserved persistent and i.i.d. shocks. Since the space of prices and shocks is finite, the strategy space is compact. Therefore, from the continuity of the objective function, the RI problem has a solution.

The state variable in the value function (5) corresponds to the prior probability of the low volatility state. Equation (7) forces the chosen joint probability distribution to be consistent with owners' prior beliefs. Without this constraint, owners could "forget" relevant information acquired in the past. Equation (8) characterizes the belief updating process. In this equation, $\mathcal{T}_{t+1}$ represents the law of motion of σ_L based on the Markov switching probabilities, while its argument is the posterior probability of the distribution of $\widehat{p}_{it}$ having low volatility at time t.

The fully-flexible information scheme imposes a challenge on how to solve (5) as the shape of $f_{it}(p_{it}, \widehat{p}_{it}|p_{t-1})$ and its implications on $\mathcal{I}(\widehat{p}_{it}, s_{it}|s_{it-1})$, has a non-linear effect on continuation values $V(m_{it+1}(\sigma_L|p_{it}))$. To tackle this issue, I build on the result given by Proposition 3 of SSM (2017).[8] The following system of non-linear equations characterizes the solution of (5) subject to equations (6) to (9).

Proposition 2 : *Solution of the dynamic RI problem*

$$m_{it}(\sigma_L|p_{it-1}) = (1 - \tau_{LH}) f_{it-1}(\sigma_L|p_{it-1}) + \tau_{HL}(1 - f_{it-1}(\sigma_L|p_{it-1})) \tag{10}$$

$$f_{it}(p_{it}|\widehat{p}_{it}, p_{it-1}) = \quad \frac{exp\left[\left(\Pi(p_{it}, \widehat{p}_{it}) + \beta V(m_{it+1}(\sigma_L|p_{it}))\right)/\lambda_i\right] f_{it}(p_{it}|p_{it-1})}{\sum_{p'_{it}} \; exp\left[\left[\Pi(p'_{it}, \widehat{p}_{it}) \; + \beta V(m_{it+1}(\sigma_L|p_{it}))\right)/\lambda_i\right] f_{it}(p'_{it}|p_{it-1})} \tag{11}$$

$$V(m_{it}(\sigma_L|p_{it})) = \quad \lambda_i E\left[\log\left(\sum_{p_{it}} (exp\left[\left(\Pi(p_{it}, \widehat{p}_{it}) + \beta V(m_{it+1}(\sigma_L|p_{it}))\right)/\lambda_i\right] f_{it}(p_{it}|p_{it-1})\right)\right] \tag{12}$$

[8] This paper argues that a dynamic RI problem consistent with (5) is equivalent to a control problem without uncertainty about $\widehat{p}_{it}$. Because of this equivalence, firm's continuation value are then a function of the *history* of prices and shocks, so the decision about the shape of the joint probability distribution does not affect $V(m_{it+1}(\sigma_L|p_{it}))$.

In Appendix 7.3, we derive these expressions. For any $\lambda_i > 0$, (10), (11) and (12) summarize the main equations that solves the problem. Equation (10) is the prior probability of being in the low volatility state as a function of the Markov transition probabilities and lagged acquired information. The expression then corresponds to the functional form of $\mathcal{T}_t$ in equation (8). The prior probability $m_{it}(\sigma_H|p_{it-1})$ is simply the complement of equation (10). These two probabilities are embedded into $g_{it}(\widehat{p}_{it}|p_{it-1})$ to force prior beliefs to be consistent with the joint probability distribution as stated in (7).

The form of the conditional probability $f_{it}(p_{it}|\widehat{p}_{it}, p_{it-1})$, i.e. the information strategy, is characterized in (11). The probability resembles the dynamic Logit formula except for the term $f_{it}(p_{it}|p_{it-1})$, which multiplies the cost-benefit ratio of choosing the price p_{it}. As $f_{it}(p_{it}|p_{it-1})$ is independent of realized shocks, it is interpreted as owner's "predisposition" to chose each $p_{it} \in \Omega_p$ without additional current information. Following SSM (2017)'s characterization, we interpret firms' predisposition as prices that are chosen with high probability (on average) across outcomes and states, i.e. $f(p_{it}|p_{it-1}) = E_{\widehat{p}_{it}}[f(p_{it}|\widehat{p}_{it}, p_{it-1})]$. The posterior probability $f_{it}(p_{it}|\widehat{p}_{it}, p_{it-1})$ is a function of λ_i as its magnitude determines the amount of information to process and, with this, the weight attached to prior probabilities. Pricing decisions are drawn from (11) reflecting the noisiness in signals, whilst being consistent with owners' idiosyncratic state-dependent beliefs. Equation (12) shows the expression for the continuation value.

Due to imperfect information about both the outcome and its time-varying distribution, there is no specific closed form for the posterior probability $f_{it}(p_t|\widehat{p}_{it}, p_{it-1})$.[9] Moreover, as the information cost non-linearly affects both the posterior probability and continuation values, it is difficult to anticipate how different values of λ_i would affect the information strategies.

SSM (2017) argued that these equations are necessary and sufficient for solving the dynamic RI problem. Given the Markovian framework, the choice rule (11) will become stationary after a finite number of periods. Therefore, the system of equations would reflect the long-run behavior of firms, so we proceed to solve it numerically.

3 Numerical solutions

Before numerically solving the model, we need to introduce further assumptions about the simplex of each variable. The computational intensity of RI models severely restricts this decision, Tutino (2013). Let $|\Omega_\epsilon| = 15$ and $|\Omega_p| = 29$ be the number of possible values that the idiosyncratic shocks ϵ_{it} and prices p_{it}, can take respectively.[10] The different values for ϵ_{it} come

[9] Starting from the same model but assuming that the underlying distribution of $\widehat{p}_{it}$ is known with certainty (i.e. a static framework), there would be a closed form expression for the posterior uncertainty. With a quadratic objective and Gaussian distributions, the model boils down to a Bayesian updating set up, where the posterior distribution of prices is equal to a weighted sum between prior beliefs and signals. Under RI, the weight attached to signals becomes the choice variable of the problem.

[10] Thus there are 870 possible result combinations from the three random variables, $f(\sigma, \epsilon, p) = 2 \times 15 \times 29$.

from a linearly equally-spaced grid ranging from $-2\sigma_H$ to $2\sigma_H$. Since $g_{it}(\widehat{p}_{it}) = m_{it}(\sigma)h(\epsilon)$, the state variable is defined as the probability of being in the low state $m_{it}(\sigma_L) \in \Delta(\Omega_\sigma)$, where $\Delta(\Omega_\sigma)$ is the belief simplex. The dimension of the belief simplex is assumed $|\Delta(\Omega_\sigma)| = 12$, where each point reflects distinct (equally spaced) values for the marginal probability of the economy being in the low volatility state. The specific details of the algorithm to solve the dynamic RI model are described in Appendix 7.4.

3.1 Calibration

The parameters in the model are: the discount rate β, switching probabilities τ_{LH} and τ_{HL}, the price elasticity η (which determines the curvature of demand γ), the volatility level in the two states $\sigma_L, \sigma_H \equiv \phi\sigma_L$, and the cost of acquiring information $\{\lambda_i\}_{i=1}^N$. Each period is a month, so consistent with an annualized risk-free rate of 4% we set the discount factor equal to $\beta = 0.9967$. The transition probabilities are assumed equal to $\tau_{LH} = 0.00882$ and $\tau_{LH} = 0.0196$, implying a quarterly probability of 2.6% of switching from the low to the high volatility state and a 94% probability of remaining in the high volatility state. These probabilities are set in order to match Bloom, Floetotto, Jaimovich, Saporta-Eksten and Terry (2018)'s estimates for the transition across uncertainty regimes in the U.S. Finally, following Nakamura and Steinsson (2008), the price elasticity of demand is assumed $\theta = 4$ (implying a 33% markup). Hence, $\gamma = -6$ as shown in Appendix 7.1.

The remaining parameters are calibrated to match different stylized facts about both micro price changes and information frictions. All the stylized facts about prices are taken from Vavra (2013), while the evidence of state-dependent information rigidities is collected from CG(2015), Section III. Based on the algorithm, we simulate an economy with $N = 5,000$ firms and $T = 1,700$ periods. In the simulations, the economy evolves naturally across states and idiosyncratic shocks, and we rule out the first 200 periods. To set the heterogeneity, we assume there are 10 distinct values for λ, which are randomly and uniformly assigned across firms, i.e. $10 \times 500 = 5,000$. Without further evidence on the cost distribution, we assume $\lambda_i \sim N(\overline{\lambda}, \sigma_\lambda^2)$ with the distribution truncated at zero. Given the calibrated parameters $\overline{\lambda}$ and σ_λ, the different values of λ are given by the 10 equidistant percentiles (from 5 to 95) of this distribution. These four parameters $\{\sigma_L, \phi, \overline{\lambda}, \sigma_\lambda^2\}$ are calibrated to match the data.

3.2 Matching moments

Table 1 shows the moments chosen from the data and their simulated counterparts from the model. $E(|\Delta p|)$ is the average magnitude of price changes (in % points) and $E(Dispersion|\sigma_i)$ is the average price-change dispersion across time when the underlying state is $i = L, H$. As these three moments are computed conditioned on a price change occurring, i.e. $|\Delta p_{i,t}| \neq 0$, they reflect the intensive margin of price changes.[11] Given the focus on unobserved volatility shocks, we assess the model's ability to replicate the price dispersion at times of low and high volatility separately.[12] We exclude "sale-type" prices by relying on the "V-shape" filter proposed by Nakamura and Steinsson (2008). Two reasons support this decision. Firstly, to provide a fair comparison with the price moments in Vavra (2013) which also remove sales, and secondly, since it has been argued that the simulated price changes of a rationally inattentive seller are consistent with sales-type (short-term) movements, Matějka (2015). The last targeted moment β_{IR} is an estimate of information rigidities as described by CG(2015). Particularly, this parameter comes from the following estimation:

$$x_{t+h} - F_t x_{t+h} = c + \beta(F_t x_{t+h} - F_{t-1} x_{t+h}) + error_t \tag{13}$$

In this equation x_{t+h} is the actual realization of the predicted variables at time $t+h$, $F_t x_{t+h}$ is the prediction for x_{t+h} using the available information up to time t, i.e. the posterior, and $F_{t-1}x_{t+h}$ is the prediction for the same variable but relying on information up to $t-1$, i.e. the prior. With a sample ranging from 1968 to 2014, CG (2015) estimate (13) for each quarter separately in order to capture the low frequency variation in the degree of information rigidity over time. The targeted moment $\beta_{IR} = 0.674$ is the average of the estimated β's across quarters and years. For completeness we also add the average confidence intervals for the estimation. By setting $h = 0$ we can repeat the same exercise and estimate (13) through the lens of our dynamic learning model and check its ability to match this parameter.[13]

We further assess the model's ability to replicate the data in two additional ways. First, by focusing on additional (non-targeted) moments from prices and secondly, through the possibility of replicating the empirical learning dynamics documented over the business cycle. Regarding the non-targeted moments in Table 1, *Frequency* and $Kurtosis(|\Delta p|)$ stand for the frequency of price reviews (fraction of prices that change) per month across firms and the kurtosis from

[11] The ability to match the distribution of the intensive margin of price changes is a relevant feature of pricing models. Midrigan (2011) documents a significant amount of heterogeneity in the (absolute) magnitude of price changes in the data. Accounting for such heterogeneity is essential as it is linked with the effectiveness of monetary policy. To reproduce this heterogeneity, we target both the mean and the variance of price changes. Moreover, we also target the price change dispersion during low and high volatility episodes. Matching the rising dispersion during high volatility episodes has also attracted the attention of pricing models. If prices become more flexible, they will absorb shocks, so the ability of monetary policy to affect the real economy would be harmed, especially during more uncertain times, see Vavra (2013).

[12] This is done by collecting Vavra (2013)'s business cycle facts and then assuming that recessions are periods of high volatility, Bloom et al. (2018)

[13] In terms of our model, with h=0 and averaging across price-setters, equation (13) becomes $\widehat{p}_t - p_t = c + \beta(p_t - p_t^p) + error_t$, where p_t^p stands for the price that the firm would set using only information collected up to $t-1$, i.e., the *predisposition* given prior beliefs. For the sake of comparison with CG(2015), we estimate (13) collapsing our simulated monthly data into quarters. This is the only moment estimated in quarterly frequency. While originally the degree of information rigidity obtained through (13) assumes $h = 4$, CG(2015) Section II.B, shows that this parameter does not change significantly for $h = 0$.

the distribution of price changes respectively. *Fraction small* is the proportion of small price changes, where a small change is defined as $|\Delta p_{i,t}| < 0.5E(|\Delta p|)$, and $Corr(Dis, Freq)$ is the correlation between the price-change dispersion and the frequency of price changes.

As shown in Table 1 the baseline dynamic RI model can simultaneously match all the targeted features of the data. The calibrated parameters are displayed in Table 2. The volatility of the price-target in the high state increases by 64%, a magnitude that is fully consistent with the estimated aggregate uncertainty parameters during episodes of economic distress, Bloom et al. (2018). The standard deviation for the attention cost $\sigma_\lambda \approx 0.08$ is considerable as it is more than half of the average cost. While the calibrations rely on a parametric assumption about the distribution of costs, they are informative as they provide a quantitative assessment about the potential dispersion of information rigidities across firms. As the ability of a RI model to match this set of moments is one of the main contributions of this paper, Appendix 7.5 provides a further discussion of the identification strategy. Although all parameters are jointly identified, we show which moments are relatively more informative of which specific parameters.

Table 1: Matched Moments and Alternative Specifications

Targeted moments	Data	Baseline
$E(\lvert\Delta p\rvert)$	0.077	0.071
$E(Dispersion\|\sigma_L)$	0.073	0.072
$E(Dispersion\|\sigma_H)$	0.090	0.090
β_{IR}	0.674	0.724
	(0.281,1.065)	
Non-Targeted moments		
$Frequency$	0.150	0.564
$Kurtosis(\lvert\Delta p\rvert)$	6.403	4.049
$Fraction\ small$	0.330	0.175
$Corr(Dis, Freq)$	0.506	0.630

Notes: All pricing moments are taken from Vavra (2013). The information rigidity parameter β_{IR} is taken from Coibion and Gorodnichenko (2015). $E(|\Delta p|)$ is the average magnitude of non-zero price changes, $E(Dispersion|\sigma_i)$ is the average price-change dispersion when the aggregate volatility state is L or H, and β_{IR} is the degree of information rigidities obtained by averaging the quarterly estimates of (13) over the sample. Frequency is the fraction of prices that change per month, $Kurtosis(|\Delta p|)$ is the kurtosis coefficient of the distribution of absolute price changes, $E(|\Delta p|)$, *Fraction small* is the percentage of small price changes, and $Corr(Dis, Freq)$ is the time series correlation between the dispersion and the frequency of price changes.

Regarding non-targeted moments, although prices remain constant in the model nearly half of the time, this is still not enough to fully replicate the frequency of price changes. This result is still relevant as the model's ability to partially replicate the degree of micro price stickiness is attained assuming imperfect information as the unique friction. Matějka (2015) showed that RI is consistent with price-setting evidence at the extensive margin. However, the intensive margin channel was muted in his analysis. The average absolute size of price changes was also calibrated by Maćkowiak and Wiederholt (2009) based on a RI model but with fully flexible prices. The results in Table 1 suggest that a dynamic version of a RI model is consistent with the evidence at the intensive margin, while generating a moderate amount of price stickiness.

Table 2: Calibrated Parameters

Parameter	Value	Description
β	0.9967	Discount Rate
γ	-6	Curvature of demand function
τ_{LH}	0.00882	Monthly transition prob.: low/high state
τ_{HL}	0.0196	Monthly transition prob.: high/high state
σ_L	0.101	Volatility in low state
ϕ	1.64	Increase in volatility in high state
$\overline{\lambda}$	0.105	Mean distribution information cost
σ_λ	0.079	Stdv distribution information cost

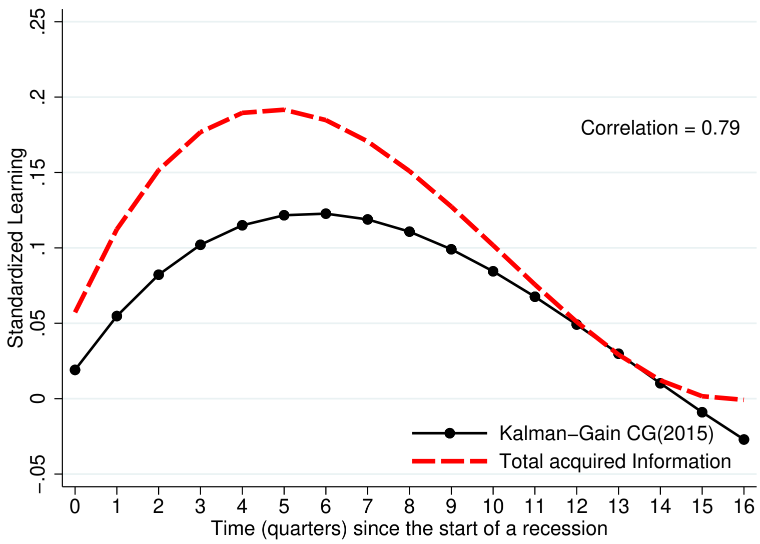
Given the two margins of price-adjustments, the baseline model is close to fully matching the positive comovement between price-change dispersion and frequency over time. Vavra (2013) argues that most price-rigidity models are at odds with this new fact and extends a canonical Ss price-setting model to reconcile its predictions with the data. Our results suggest that this feature of the data is also consistent with our setup. Section 4 provides further intuition on the mechanism by which the model can rationalize this correlation.

As information acquisition is fully endogenous in the model, we expect that when the economy enters into a high volatility state, firms would decide to gather more information before setting prices. While based on the RI literature we anticipate that this will happen, Maćkowiak and Wiederholt (2009), it is less obvious that the learning dynamics will actually resemble the response of information rigidities over the business cycle. Using the calibrated parameters in Table 2, we estimate:

$$\overline{\kappa}_t = \alpha + \sum_{j=0}^{J} \phi_j I_{t-j}^{REC} + error_t \tag{14}$$

Where $\overline{\kappa}_t \equiv \sum_i^N \kappa_{it}$ is the average total acquired information across firms at time t, and I_{t-j}^{REC} is a dummy variable equal to 1 which marks the first quarter after which the economy enters into a high volatility state. Equation (14) again follows CG(2015). By changing j we can back-out a sequence of ϕ_j that are interpreted as an impulse response of average acquired attention to a high volatility state. The same equation is estimated by CG(2015) but instead of $\overline{\kappa}_t$ they use their estimated measure for information rigidities β_t obtained through (13). The final external validation of our baseline model is to assess whether it can replicate the learning dynamics documented by these authors. To provide a fair comparison and given the one-to-one mapping between β_t and the weight on current signals $\widehat{G} = 1/(1+\widehat{\beta})$, i.e. the Kalman Gain discussed by CG(2015), we compare the IRFs of (14) against the ϕ's obtained through the same equation but replacing $\overline{\kappa}_t$ with $\widehat{G}$. For this we use CG(2015)'s original data and aggregate our simulated $\overline{\kappa}_t$ in quarterly frequency to mirroring their data. We standardized the two dependent variables to account for the different units of the IRFs, where the first is measured in information bits and the second is a relative weight. Thus, we interpret the responses in standard deviation units. This is shown in Figure 1.

Figure 1: Information acquisition over the business cycle



Notes: The red line of the figure shows the response of the standardized $\overline{\kappa}_t$ after estimating (14). The black line is the response of the standardized $\widehat{G} = 1/(1+\widehat{\beta})$ using CG(2015)'s original data. In this latter case, we replace $\overline{\kappa}_t$ for $\widehat{G}$ in (14). Consistent with CG(2015) we smooth the response of both coefficients by fitting a polynomial distributed lag model where the order of the polynomial is 3 and with total lags J=20.

As shown by Figure 1, acquired information increases after the economy enters into the less predictable state, i.e., $t = 0$, in a matter consistent with its empirical counterpart. In fact, the response upon impact is relatively similar between the two models, with a slightly higher level for the endogenous attention specification. As in the data the adjustment is, however, sluggish meaning that firms can not immediately recognize that the economy is in a different state. The specific features that cause this delayed learning patterns are discussed extensively in section 4. Interestingly, although the amplification of the dynamic RI model is higher, the two models reach their maximum level at around the fifth quarter. The correlation of 0.79 of the responses is reassuring as it reflects that quantitatively the two models behave similarly after a second-moment shock. By studying the implications of a decision maker, a price-setter, who faces costly entropy reductions within a dynamic learning setting, we were are able to rationalize the attention dynamics over the business cycle. This reinforces the importance of allowing for inattention at the firm level to better understand decisions.

3.3 Information and pricing strategies

Let us now focus on the implications of acquired information on price setting given the different costs. At each t owners choose their information strategies, which will support their pricing decisions. Uncertainty about the current distribution of $\widehat{p}_{it}$ implies that the decision about the amount of information to collect is linked with the prior beliefs about each state. Due to the cost heterogeneity and for expository reasons we focus on three type of firms, specifically those facing low, medium, and high information costs. In particular, we focus on firms with $\lambda_1 < \lambda_5 < \lambda_{10}$ taken from the cost distribution.[14]

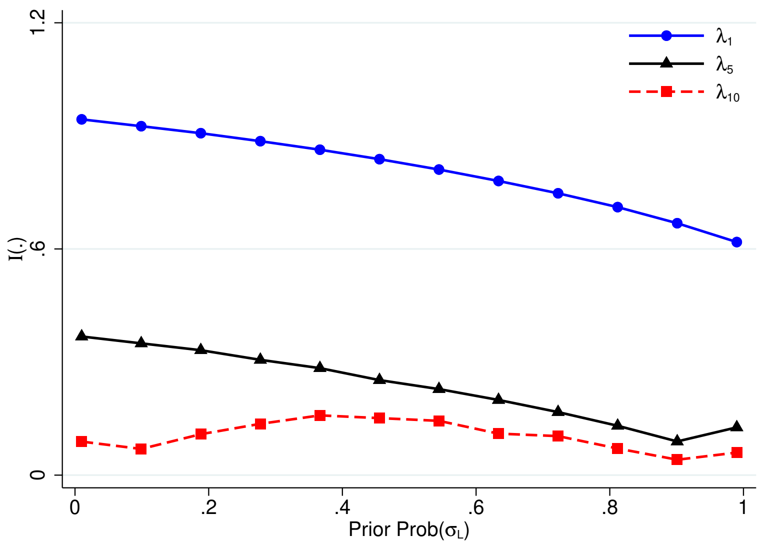
Figure 2 shows $\mathcal{I}(p_{it}, \widehat{p}_{it}|p_{it-1})$, as a function of distinct prior probabilities assigned to the low volatility state for these three firms. As the prior converges to one, all firms choose to acquire less information as a response to the idiosyncratic shocks being more predictable. However, as the prior approaches zero, firms face an information trade-off. On the one hand, they are almost certain that the economy is in the high volatility state which discourages acquiring costly information on the state of the economy, while on the other hand, gathering additional information is useful for responding to the less predictable idiosyncratic shock. Intuitively, which of the two mechanisms dominates is driven by the magnitude of λ. The results support this intuition. As the perceived predictability of the unobserved target decreases, it becomes increasingly optimal to acquire more information. However this is only true for low and medium costs. The information profile for higher cost firms exhibits an inverted U-pattern indicating that the increased certainty about being in the high volatility state crowds-out the final amount of information acquired.

Let us turn to the chosen information strategies $f_{it}(p_{it}, \sigma_t, \epsilon_{it}|p_{it-1})$ for the three cost levels. To minimize conditional variance (equation (4)) firms design their optimal information and pricing strategies conditioned on the costs they face. Information costs not only affect the search quality, they also force owners to form a probabilistic conjecture about the likelihood of extreme realizations of the target, which ultimately affects their pricing decisions.

For expository reasons, we compute strategies when price setters believe there is a 75%, 50% and 25% probability of being in the low volatility state. As the joint probability distribution depends on three random variables, we calculate the "predisposition" $f_{it}(p_{it}|p_{it-1}) = \sum_{\sigma_t}\sum_{\epsilon_{it}} f_{it}(p_{it}, \sigma_t, \epsilon_t|p_{it-1})$, where prior beliefs about σ_t are embedded in p_{it-1}. This is presented in the left column of Figure 3, starting from the highest cost. With new information about $\widehat{p}_{it}$, firms update their beliefs to form the posterior $f_{it}(p_{it}|\widehat{p}_{it}, p_{it-1})$ from which the optimal

[14] Given the calibrated $\overline{\lambda}$ and σ_λ of the truncated Normal distribution for the information costs, we choose the middle cost λ_5 to resemble the average firm. In particular, according to the calibration, $\lambda_5 = 0.1055$ and $\overline{\lambda} = 0.1046$ as reported in Table 2.

Figure 2: Total Information



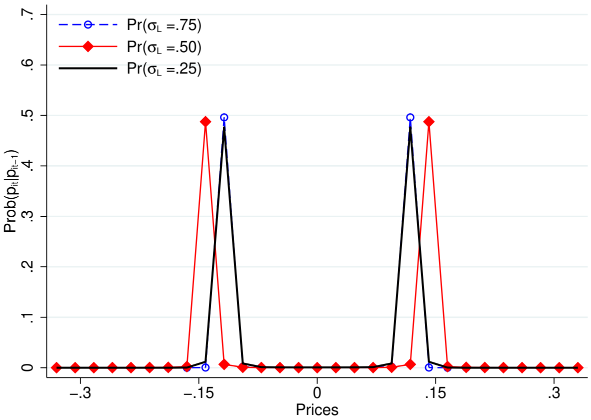
Notes: the figure presents total acquired information with respect to prior probabilities attached to the economy being in the low volatility state. The relationship is shown for three different information-cost values, $\lambda_1 < \lambda_5 < \lambda_{10}$.

price is finally drawn. This is shown in the right panel of the figure. In all cases we assume a given constant realization for $\widehat{p}_{it}$ drawn from the low volatility distribution. The optimal price is represented by the vertical solid line.

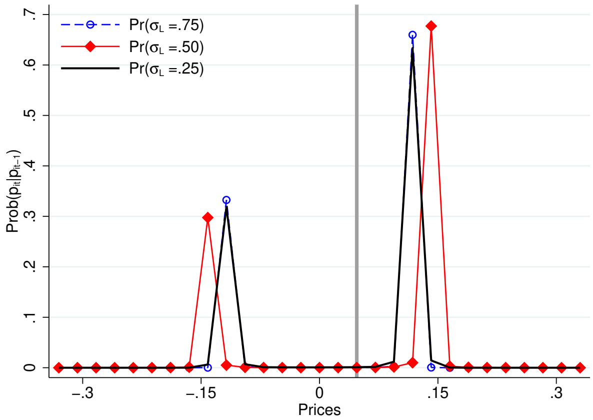
The results show that a rational inattentive price-setter is prone to choose from a smaller price set relative to the full array of possible prices. This feature is labelled as the "optimal consideration set" by the RI literature, Caplin, Dean and Leahy (2019). From the prior probabilities, it is clear that the range of possible values for p_{it} decreases with the magnitude of the information cost. For λ_{10} the predisposition $f_{it}(p_{it}|p_{it-1})$ degenerates into two possible prices. Following equation (11), the decision is to assign almost zero probability to choosing a price different from these two. The chosen magnitude of the two possible prices is, however, relevant as owners are trying to reduce their probability of making large mistakes on average to the utmost. When the uncertainty about the correct distribution is maximized, i.e., $Prob(\sigma_L) = 0.5$, the λ_{10} firm responds by choosing between more extreme price values. In line with Figure 2, as a response to the perceived higher probability for the σ_H state, the firm chooses to acquire less information and return to its previous price plan where it discriminates between two possible prices closer to the mean.[15]

[15]Despite that for some values of λ_i, the predisposition concentrates almost all the likelihood in a subset of prices, i.e., it degenerates, numerically we confirm that the probability of choosing each possible p_{it} in Ω_p is always strictly greater than zero for any value of λ and $g_{it} \in \Delta(\Omega_{\widehat{p}})$. Hence, the solution is indeed interior.

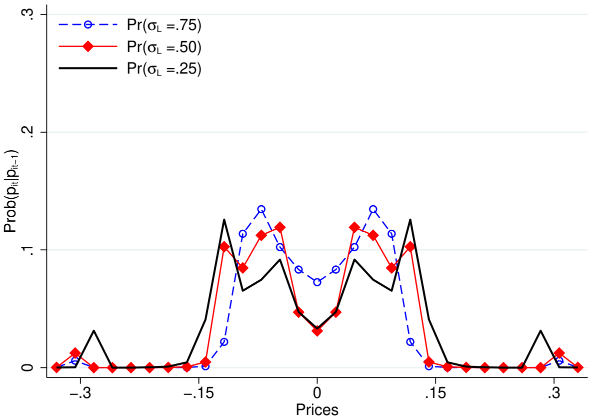
Figure 3: Firm's predisposition $f(p_{it}|p_{it-1})$ and conditional probability $f(p_{it}|\widehat{p}_{it}, p_{it-1})$



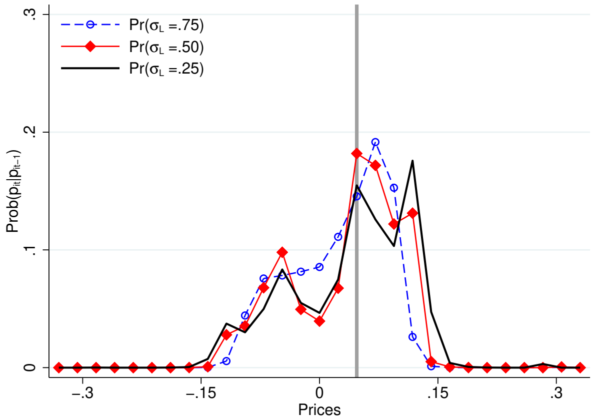
(a) Prior Probability λ_{10}



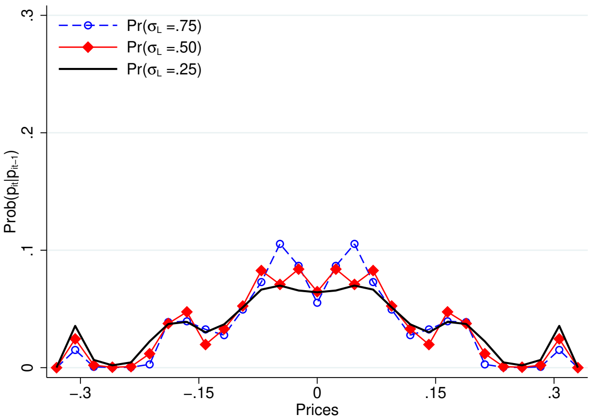
(b) Posterior Probability λ_{10}



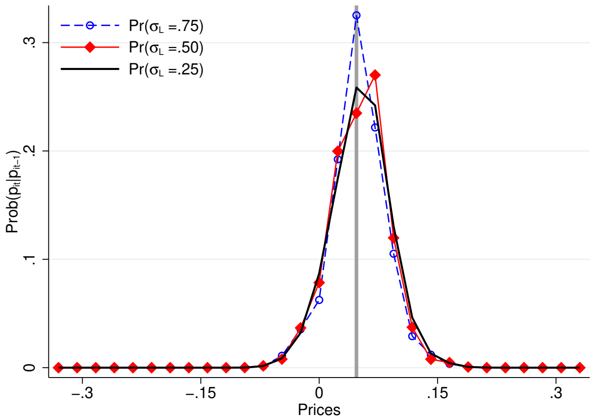
(c) Prior Probability λ_5



(d) Posterior Probability λ_5



(e) Prior Probability λ_1



(f) Posterior Probability λ_1

Since this "degenerate" pricing strategy is common across firms with high information costs, the model resembles price stickiness. Moreover, while the pricing strategies look similar for $Prob(\sigma_L) = 0.25$ and 0.75, in the former case the firm chooses to attach some small probability around the price with the highest probability. This increases the frequency of price adjustments as the economy moves towards the less predictable state, a feature consistent with the data.

Cheaper information allows the owner to more precisely distinguish both the realization of the target-price and its underlying distribution. The λ_5 firms are inclined to choose among a wider set of prices but still assign almost zero probability to extreme realizations of $\widehat{p}_{it}$, with the exception of $Prob(\sigma_L) = 0.25$. The predisposition for λ_1 resembles a normal distribution where, instead of concentrating all the probability mass in a couple or a small subset of prices, the firm allocates the prior probability not to rule out any potential price ex-ante. The shape of the predisposition allows them to react immediately as new (and very precise) information arrives, distributing the posterior probability close to the true realization. This is shown by the lowest graph in the right column of Figure 3. Around 28% of time the λ_1 firm correctly set $p_{it} = \widehat{p}_{it}$ for the three probability scenarios. When the probability of σ_L decreases and firms collect more information, the distribution changes to attaching more probability to extreme realizations, in exchange with reducing the probability of setting prices closer to the mean. Even if the firm wrongly attaches a high probability to the more volatile distribution, it is still able to closely track the optimal price.

As noticed, pricing decisions are error-prone under RI. Given the significant dispersion of the information costs, we can quantify the expected per period profit loss from these mistakes across the distribution of firms. In particular, we can compute the expected static profit loss under RI with respect to the Full-Information (FI) profit, i.e., the frictionless equilibrium where firms observe $\widehat{p}_{it}$ without noise. In our model, this is equivalent to assuming that $\lambda_i = 0$ for all firms. For the average firm, pricing mistakes from imperfect information account for a decrease of 5.7% in profits relative to the FI benchmark. Firms facing the lowest λ_1 and the highest cost λ_{10} of information reduce their profits by 0.9% and 10.4%, respectively, in comparison with the frictionless scenario. We find numbers in this range reasonable as they do not force firms to face disproportionally large profit losses from inattention while generating a significant degree of sluggishness in learning as described in the following Section.

Despite the quality of the signal, any firm can have misperceived beliefs about the correct distribution. If the economy switches to the low volatility state after being in the high state for several periods, even a highly informed owner may continue setting prices thinking that the economy is still in the less predictable state. It would be enough that the current realization for $\widehat{p}_{it}$ is close to the mean in order to prevent them from noticing any differences. The rate by which firms uncover a potential new state, which is linked to their information and pricing strategies, is what brings most of the interesting insights to the problem.

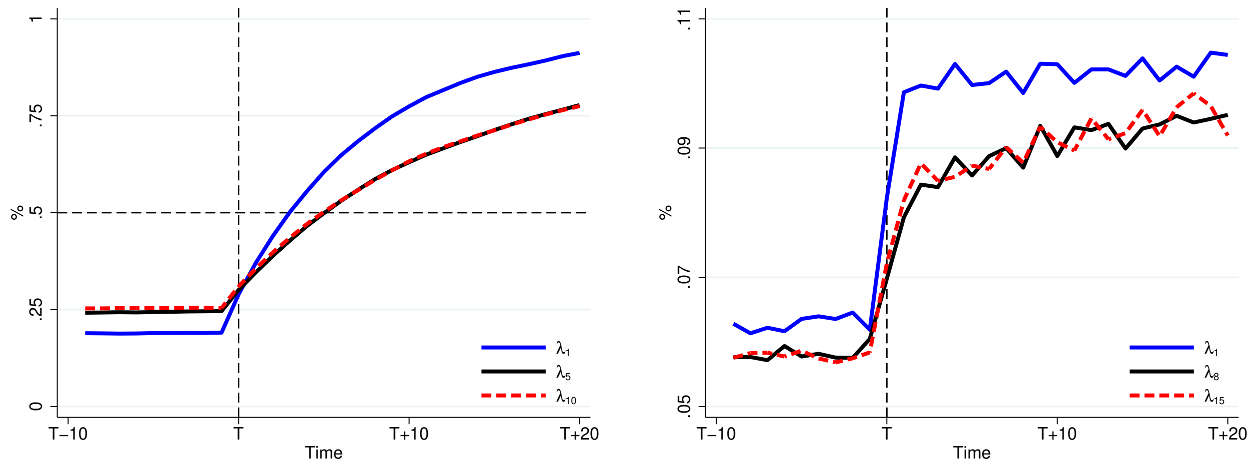
4 Delayed learning dynamics

We will now study the transition dynamics of the model by simulating an exogenous state change. At the initial time $T = 0$ we assume that firms know the correct state with certainty, after which they start to update their beliefs about the current state. We assume that the economy remains in the low volatility state for 100 periods, and then switches to the high state at time T. Keeping the assumed transition of states constant, we simulate 50 economies with 5,000 firms. As in the calibrations, we allocate the 10 different information costs uniformly across these 5,000 firms. The results are then the averages of the variables across economies at each point in time.

4.1 Persistent beliefs at the firm level

We start by presenting the results at the firm level by focusing on the same three groups of firms, λ_1, λ_5, and λ_{10}, as in the previous section. Figure 4 shows the evolution of posterior beliefs about the high volatility state $f_{it}(\sigma_H|p_t, p_{it-1})$ and the average price-change dispersion $E(Dispersion|\sigma_j)$ with $j = L, H$, for the three type of firms after the aggregate state change (vertical dotted line).

Figure 4: Firm's posterior beliefs and average price-change dispersion



(a) Posterior $f(\sigma_H|p_{it}, p_{it-1})$ - Low $\rightarrow$ High (b) $E(Dispersion)$ - Low $\rightarrow$ High

Notes: the vertical dotted black lines represent the moment when the economy switches from the low to the high volatility state. The left figure presents the evolution of posterior beliefs for the three firm types. The right panel shows the average price-adjustment dispersion across firm type.

When information is cheap, owners can more easily notice a state change and respond to the lower predictability of the target by acquiring more information. However, even firms with the lowest information cost λ_1 do not immediately notice the new distribution causing a sluggish reaction in the revision of their posterior beliefs, affecting the amount of acquired information. Hence, imperfect information about persistent volatility states endogenously generates persistence in beliefs. The reaction of firms facing a higher information cost, λ_5 and λ_{10}, is even more sluggish. As the higher costs lead firms to focus their attention on a smaller consideration set, this delays the rate by which these firms recognize more extreme realizations of $\widehat{p}_{it}$. Following the simulated exercise, while firms with the lowest cost still need approximately four months to start attaching more probability to being in σ_H, firms with middle or higher costs need seven months. The prediction about heterogeneous beliefs about a common variable coexisting between firms during the same time is also supported by the empirical evidence, see Kumar, Afrouzi, Coibion and Gorodnichenko (2015).[16]

Active learning within this dynamic setting generates disagreement about the correct underlying distribution for $\widehat{p}_{it}$, which affects both pricing and posterior information decisions. The evolution of price dispersion is presented in the right panel of Figure 4. In the new volatility state, the price dispersion increases reflecting the intrinsically less predictable price-target. However, with the new state, the price dispersion displays an upward trend reflecting that this subset of firms keep setting prices based on outdated beliefs about the true price distribution.

4.2 Aggregate evolution

Let us now turn to aggregate pricing decisions. Figure 5 presents the evolution of the price-change dispersion (given by the standard deviation of Δp_{it}) and, in the secondary axis, the evolution of the frequency of price changes for the same assumed transition. As the economy change states, both the price-change dispersion and the frequency of price changes commove positively consistent with the data. While around half of the time prices are not adjusted in the model, we get an average increase of 5.5% during the high volatility state for the extensive margin.

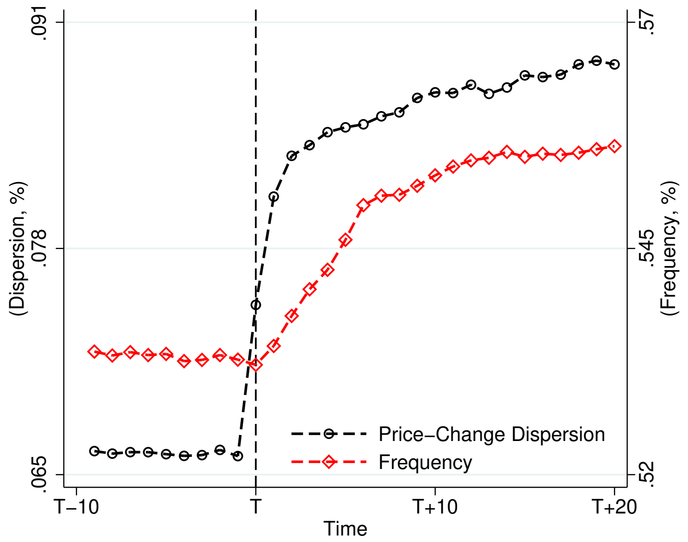
As previously discussed, while a growing proportion of firms are noticing any new state, the rest will keep setting prices as if the state has remained constant for some additional months, affecting the overall price-change dispersion. More informed firms also adjust their prices more frequently depending on the perceived predictability of the state.

Under full information, the volatility of $\widehat{p}_{it}$ is $\sigma_L \approx 0.1$ and $\sigma_H \approx 0.16$ (Table 2). Given the chosen learning and pricing strategies, the presence of Rational Inattentive price-setters diminishes the price dispersion relative to a scenario where acquiring information is not costly. Moreover, and according to the calibrated parameters, while volatility increases by 64% in the high state, the effective price dispersion increase is around half of it, accounting for 29%. Hence the implications of firms endogenously choosing not to waste attention on noticing any extreme

[16] Although this paper documents the presence of time-varying heterogeneous beliefs about inflation, we see this as a valid proxy for the beliefs about an aggregate price index such as $\widehat{p}_{it}$.

realization of the target-price ends up affecting not only the level of dispersion but also its underlying dynamics.

Figure 5: Aggregate Evolution



Notes: the figure presents the time series evolution of price-change dispersion and the frequency of price changes (secondary axis). The dotted vertical lines show the moment when the economy switches to the low or high volatility state respectively.

Through the results, we can map the implications of the proposed information-driven mechanism for price stability. This can have interesting consequences for policy design. In particular, the scope by which policies can effectively reduce price instabilities can be very different depending on agents' beliefs about the overall state of the economy. For example, let us assume that agents are confident that the high volatility state is indeed the current state of the economy. The two overlapping distributions for $\widehat{p}_{it}$ and the pricing strategies when the prior beliefs of being in σ_L are low, are possible enough to force a situation where prices are persistently more volatile over time independently of the actual state.

5 Alternative specifications

The main ingredients of the model are dynamic information, heterogeneous information costs, and idiosyncratic optimal prices $\widehat{p}_{it}$. In this section, we explore the model's ability to replicate the aforementioned stylized facts after shutting down each of these channels in turn. As the results of our baseline setup are consistent with both empirical evidence on micro prices and the dynamics of acquired information, through this exercise we can highlight the specific elements that account for each of these results. We start by presenting each alternative version separately.

5.1 A Static Learning Model

Although the price-setting problem of the baseline model is fairly stylized, allowing for $\widehat{p}_{it}$ following a mixture of normal distribution is already enough to deliver non closed-form solutions. However, the model has a straightforward static counterpart version, which is relevant as it delivers analytical implications.

This alternative version of the model is labelled as *Static Learning*. We keep the same structure for $\widehat{p}_{it} = \sigma_t \epsilon_{it}$ as before, but instead we assume full information about its current distribution. Hence, firms acquire costly information to *only* track the realizations of $\widehat{p}_{it}$. The problem becomes very tractable as it turns into a standard static RI problem with a quadratic objective (1) and $\widehat{p}_{it} \sim N(0, \sigma_j^2)$, given the known aggregate state $j = L, H$. As discussed by Maćkowiak et al. (2018), under these two assumptions each price-setter learns through normally distributed signals $s_{it} = \widehat{p}_{it} + \eta_{it}$ with $\eta_t \sim N(0, \sigma_{i\eta}^2)$. Under active learning, the precision of the signal $\sigma_{i\eta}^{-2}$ is chosen by each firm.

The price-setter chooses the amount of attention κ_{it} to pay subject to the information costs $\lambda_i > 0$. By solving the firm's problem, we get a closed-form expression for total acquired information:

$$\kappa_{it}^* = \begin{cases} \frac{1}{2}\log_2\left(\frac{2\gamma \ln(2)\sigma_t^2}{\lambda_i}\right) & \text{if} \quad \left(\frac{2\gamma \ln(2)\sigma_t^2}{\lambda_i}\right) > 1, \quad \text{with} \quad t = L, H \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Given κ_{it}^* the optimal price set by firms is equal to $p_{it} = (1 - 2^{-2\kappa_{it}^*})(\widehat{p}_{it} + \eta_{it})$, see Wiederholt (2010). Consistent with our baseline scenario, and in line with equation (15), price-setters choose to acquire more information the minute they realize the economy is in the more volatile state or if they face cheaper information costs.

5.2 Homogeneous information costs

In this version, we assess the role that heterogeneous costs play in the dynamic learning model. We again replicate our baseline model, with the only difference that now all firms are ex-ante identical in terms of their costs λ, i.e., $\sigma_\lambda = 0$.

5.3 Common price-target

This alternative specification rules out the presence of idiosyncratic shocks i.e. $\widehat{p}_{it} = \widehat{p}_t = \sigma_t \epsilon_t$ in the baseline model. While keeping costs heterogeneous, firms now collect information about both the distribution and realization of the *same* target-price. This assumption creates further incentives for owners to anticipate what other firms are doing. Including this feature would severely complicate the model as now owners should form higher-order beliefs about the aggregate price within a fully-flexible setting.[17] To rule out this possibility, and for the sake of tractability, we assume that owners operate in segmented markets. All the remaining assumptions are the same.

5.4 Keeping "sales-like" movements

Related to the assumed drivers of the volatility of $\widehat{p}_{it}$ in the model, and despite the theoretical predictions of the RI literature, the model's ability to quantitatively match the targeted moments could be challenged by the application of the sales filter. Therefore, we also propose and calibrate a "No Filter" version of the baseline model.

5.5 Implications of alternative models

For each alternative version of the model, we repeat the same calibration strategy to search for the parameters that minimize the distance between the simulated moments and their empirical counterparts. To contrast with our baseline specification, we again rely on the Nakamura and Steinsson (2008)'s "v-shape filter" with the obvious exception of the No Filter version. The simulated moments are presented in Table 3.

5.5.1 The relevance of dynamic information

To provide a fair comparison with respect to the static learning setup, we choose not to evaluate this model based on its capacity to replicate β_{IR}. The original approach of CG(2015) was to introduce a test whereby studying the relationship between ex-post forecast errors and their ex-ante forecast revisions (13), we can map the underlying degree of information rigidities. However, the learning structure and particularly the test were designed for predictions about a persistent stationary variable, which is not the case in the static version. Hence, besides the three original targeted moments used in the baseline setup, we choose to target the fraction of small price adjustments instead of β_{IR}. This latter assumption does not prevent us from

[17] As discussed by Hellwig and Veldkamp (2009), when a large number of agents play a game with strategic complementarity, the choice of information exhibits complementarity as well, creating an incentive for agents to learn what others are learning. Extending our dynamic learning model to allow for strategic complementarity is something beyond the motivation of this paper.

Table 3: Matched Moments and Alternative Specifications

Targeted moments	Data	Baseline	Static	$\lambda_i = \overline{\lambda}$	$\widehat{p}_{it} = \widehat{p}_t$	No Filter
$E(\|\Delta p\|)$	0.077	0.071	0.082	0.062	0.069	0.073
$E(Dispersion\|\sigma_L)$	0.073	0.072	0.128	0.069	0.052	0.077
$E(Dispersion\|\sigma_H)$	0.090	0.090	0.133	0.086	0.051	0.097
β_{IR}	0.674	0.724	-	0.657	0.653	0.734
	(0.281,1.065)					
Non-Targeted moments						
Frequency	0.150	0.564	0.831	0.742	0.577	0.781
$Kurtosis(\|\Delta p\|)$	6.403	4.049	4.863	3.919	3.921	4.199
Fraction small	0.330	0.175	0.272	0.211	0.163	0.219
$Corr(Dis, Freq)$	0.506	0.630	0.036	0.738	-0.067	0.578

Notes: All pricing moments are taken from Vavra (2013). The information rigidity parameter β_{IR} is taken from Coibion and Gorodnichenko (2015). $E(|\Delta p|)$ is the average magnitude of non-zero price changes, $E(Dispersion|\sigma_i)$ is the average price-change dispersion when the aggregate volatility state is L or H, and β_{IR} is the degree of information rigidities obtained by averaging the quarterly estimates of (13) over the sample. Frequency is the fraction of prices that change per month, $Kurtosis(|\Delta p|)$ is the kurtosis coefficient of the distribution of absolute price changes, $E(|\Delta p|)$, *Fraction small* is the percentage of small price changes, and $Corr(Dis, Freq)$ is the time series correlation between the dispersion and the frequency of price changes.

assessing the static model on its ability to match the dynamic features of attention over the business cycle. This is discussed in section 7.6 of the Appendix.

The static model fails to match the four targeted price moments. Although the model is close to matching the average magnitude of price-adjustments and the fraction of prices revised by a small magnitude, it poorly fails to match the average dispersion at both low and high volatility states. Moreover, the simulated price dispersion is not only higher relative to the data, but it is also very similar in the two states. As the problem of tracking the realization of the target price is relatively more straightforward and since firms immediately adjust κ_{it} as the economy evolves through the different states, there are no significant discrepancies in the two volatility levels.

Related to non-targeted moments, the static model is again inconsistent with the data. Given $p_{it} = (1 - 2^{-2\kappa^*_{it}})(\widehat{p}_{it} + \eta_{it})$, the randomness in the signal s_{it} leads the optimal price to change constantly through time. Given the assumed discretization for p_{it} and the sales filter, around 83% of firms decide to adjust their prices between two consecutive periods.[18] Although the extensive margin of price adjustments is not also well targeted by the baseline model, it still has implications for the correlation between the price dispersion and the frequency of adjustment. Given p_{it} and (15), prices change at the same rate independently of the aggregate volatility state leading the correlation between the two margins being almost zero.

[18] As expected, with a finer grid for p_{it} practically *all* firms adjust their prices at every time t despite the sales filter. We run additional simulations with more grid points to confirm this intuition.

5.5.2 The relevance of heterogeneity

Since the homogeneous cost version of the model has only three parameters, we target $E(|\Delta p|)$, $E(Dispersion|\sigma_L)$, and β_{IR}. Despite the optimal price being different, sharing the same λ and given the process for $\widehat{p}_{it}$, firms will have a common pricing strategy.

As shown by the fifth column in Table 3, the model's fit with homogeneous cost is not significantly worse than the baseline model. This is informative as it suggests that the extent to which we can replicate the data is not entirely driven by a vast proportion of firms making large profit losses from inattention. Adding heterogeneous λ's allows for the coexistence of small and large price adjustment through different pricing strategies, which affect the mean and the dispersion of the price-adjustment distribution. While the version with homogeneous costs cannot replicate these three features of prices as close as the baseline model, it is very accurate in matching the degree of information rigidity in the data. The frequency of price adjustments is also greater under common information relative to the baseline scenario, as well as the correlation between the two price margins.

The model where the price-target is common is successful in matching the average adjustment magnitude and the degree of information rigidity β_{IR}. However, the price-change dispersion reaction is not consistent with the expected effects of a volatility shock. As firms are tracking the exact same price, and since they choose to collect more information depending on the target's perceived predictability, there are no many discrepancies in the magnitude by which owners adjust their prices, independently of the state. Hence, price dispersion is unresponsive over the cycle. This ultimately rules out the possibility of matching the positive correlation between dispersion and the frequency of price adjustments. As observed in the last row of Table 3 the correlation is almost zero in this case.

Besides these moments, and to provide a complete comparison with the baseline specification, Section 7.6 in the Appendix discusses the evolution of endogenous attention over the business cycle for all the alternative models. Consistent with the aforementioned results (and with the exception of the common price version), none of these models is able to match the dynamic attention response over the business cycle found in the data.

5.5.3 Robustness to sales patterns

According to the results in the last column of Table 3, the "No Filter" version of the model can also match the pricing data without differing much from its baseline counterpart. As expected, the frequency of price changes increases by 38% approximately without the filter.[19] Moreover, the positive correlation between frequency and dispersion still arises in this model.

[19] According to Nakamura and Steinsson (2008), see Table IX, the frequency of price revisions (excluding product substitutions) rises from 15.3 to 19.4 when sales are not filtered out. This implies a 27% increase approximately.

This result reinforces the discussed implications of delayed learning to match this feature of the data without obscuring its relevance based on how the simulated data was treated.

5.6 First and second-moment shocks

In the baseline model, the optimal price is only affected by a second-moment shock. We deliberately allow for just one shock to cleanly isolate the novel transmission channel of dynamic learning into firms' pricing decisions. However, recessions are episodes characterized by a negative first-moment shock *combined* with a positive second-moment shock, which affects both macro and micro-processes, Bloom et al. (2018). This could raise some doubts about the extent to which our results hold when we allow for these two shocks to affect the target-price simultaneously. This section proposes an extended version of our model to account for these business cycle features.

We assume that after the economy enters into a high volatility state, the mean of $\widehat{p}_{it}$ drops at the same time, resembling a recession. We redefine the process for the optimal price as:

$$\widehat{p}_{it} = \mu_t + \sigma_t \epsilon_{it} \tag{16}$$

$$\mu_t = \begin{cases} \mu_L, & \text{if} \quad \sigma_t = \sigma_L \\ \mu_H, & \text{if} \quad \sigma_t = \sigma_H \end{cases} \tag{17}$$

Where $\mu_H < \mu_L$ and, as described in Section 2.1, the volatility in the high state is $\sigma_H = \phi\sigma_L$, $\epsilon_{it} \sim N(0,1)$, and the transition probabilities are still τ_{LH} and τ_{HL}.[20]

The learning incentives are, however, potentially different in this scenario relative to our baseline model. In the original model, an increase in the perceived dispersion of signals was the only piece of suggestive evidence of a change of state. Adding the first-moment shock also centers the received signals, and consequently the actions, around two different values. This could bring two possible counteracting effects for the decision to acquire information. On the one hand, a drop in the average of signals could boost learning by suggesting that the economy is now in a more volatile state. On the other hand, by anticipating that signals are now centered around two different values, and given the cost of acquiring information λ_i, firms could be willing to modify their learning strategies towards less attention in response to a possibly easier problem. Which of these two effects will dominate is intrinsically related to the model's parameterization.

Therefore, and for the sake of comparison, we solve this extended model using the same parameters as in Table 2. The idea is not to distort any parameter, in particular the ones

[20] The definition in equation (16) does not changes the interpretation for the target-price as a stochastic deviation from a trend. We can assume that $\log(P^*_{it}) = \mu_t + t + \sigma_t \epsilon_t$ where, as before, the trend t is known.

related to the average cost of information λ_i and its dispersion across firms σ_λ.[21] The drop in μ_t during a recession is calibrated to be consistent with the empirical evidence. Using a VAR model of factor prices for the US, Bloom (2009) shows that an uncertainty shock brings a reduction in overall prices of 0.5% approximately. Therefore, we set $\mu_H = 0.995\mu_L$ in the simulations.

Table 4 and Figure 8 in Appendix 7.7, shows the pricing and information moments along with the impulse response of total acquired information in the two shocks model. Notably, the capacity to generate the positive correlation between dispersion and frequency is still present. Information acquisition is again counter-cyclical when we allow the volatility shock to be supplemented with a first-moment shock. As expected, the learning dynamics over the business cycle are different. The presence of the time-varying mean helps firms to recognize a new state faster and therefore boosts the acquisition of information compared to the baseline model. The dynamic correlation of both responses with the data is high enough not to compromise the main implications of the pricing model already discussed. If we shut down the second moment and instead allow for just a first-moment shock to the target price, information acquisition remains relatively constant over time. We refer to Appendix 7.7 for more details about this specification.

5.7 Further alternative specifications

As discussed, our price-setting dynamic learning model can match non-trivial moments for the micro price evidence, typically addressed by state-dependent pricing models, while being consistent with the empirical features of time-varying information acquisition. This latter feature is not captured by menu-cost models, which are by construction muted when it comes to generating interesting predictions about state-dependence in information rigidities.

There are other papers that, as in our case, study price-adjustment decisions by allowing for information processing costs without assuming further frictions. Costain and Nakov (2015) and Costain and Nakov (2019) study price stickiness subject to control costs which delivers error-prone decisions. By choosing the level of control over pricing decisions, the firm limits the randomness, allowing it to set prices with higher precision. In particular, this precision is measured in terms of relative entropy between the distribution of actions (from which the price is drawn) and an exogenous default distribution, which is assumed to be uniform. The pricing adjustment mechanism is static as it takes the form of a multinomial logit where each possible decision, due to the uniform prior, is weighted equally. The rigidity of this structure force learning and pricing decisions to be independent of the state of the economy, preventing the model from accounting for the time-varying frequency of price adjustments and information

[21] As this is intrinsically a new model with different incentives if we allow the parameters to vary it would be difficult to assess if the responses are driven by adding a time-varying mean or because a change in the cost of information alters the cost-benefit analysis from which firms designs their learning strategies.

acquisition decisions. As discussed, these are two features that are consistent with our dynamic learning model.

A further extension, not explored in this paper, is to move beyond quadratic objective functions allowing for more realistic profit functions. This could push the results to resemble the price-stickiness observed in the data, as shown by Matějka (2015). However, extending the solution to dynamic models with non-quadratic objectives is even more challenging due to computational costs and lack of tractability. This is partly since we need to solve *simultaneously* the information acquisition strategy and the decision-making tasks within different prior beliefs about a persistent variable. Through the quadratic framework with Gaussian signals, we can separate these two decisions making a dynamic model more tractable, Tutino (2013), but still being very sensitive to small modifications in the problem, Jung et al. (2019). In the dynamic model presented in this paper, although price-setters learn about a mixture of normal distributions, the quadratic objective is necessary to keep the problem fairly manageable with the advantage of being effective in matching the aforementioned stylized facts. As shown in Section 5.5.1, a more stylized environment with analytical results comes at the expense of delivering non-consistent features of the data.

6 Conclusion

This paper addresses price-setting decisions under dynamic imperfect information acquisition. Rational inattentive price-setters collect information about an unobserved target-price before setting prices. Costly information serves two purposes: helping to determine the realization of the variable along with shedding light about the distribution that generated the target. The unobserved distribution is time-varying to allow for different states of the economy where aggregate volatility can rise. Information is dynamic and fully flexible as owners choose the amount of information to acquire as well as how they want to learn about outcomes. This mechanism generates persistence in beliefs, which is crucial to match distinct features of the micro price data and the empirical evidence on state-dependent attention.

While imperfect information is enough to match the dynamic features of the data and can generate price stickiness, it is still not enough to exactly fit the fact that prices remain constant for several periods. Although the model endogenously generates some degree of price stickiness, the simulated duration is shorter than the data. Adding further frictions, such as price rigidities within the described dynamic learning structure, emerges as a natural extension of this paper, in the line of Woodford (2009) or Stevens (2019). The combination of menu-costs with heterogeneous persistent beliefs would presumably amplify the documented effects on price dispersion as the economy moves across different volatility states.

As the algorithm to solve the model does not depend on any specific objective function or on a particular parametric distribution for the unobserved shocks, it can be naturally extended to alternative settings beyond price-setting decisions. Concerning dynamic learning, it is relevant to more deeply explore the consequences of endogenous asymmetric learning rates over different states of the economy. While Van Nieuwerburgh and Veldkamp (2006) studied asymmetric responses due to imperfect information, there is no additional evidence in the context of costly entropy reduction, where heterogeneous learning rates arise endogenously due to agents' private efforts.

This paper's main motivation was to assess the time-varying implications of costly information and its joint relationship with overall price stability and state-dependent learning. The fact that a model with optimal allocation of costly attention is consistent with two different strands of the literature is important for macroeconomic policy modeling and for further research on the dynamic consequences of information rigidities.

References

Afrouzi, Hassan, "Strategic Inattention, Inflation Dynamics and the Non-Neutrality of Money," 2018.

Alvarez, Fernando E and Francesco Lippi, "A note on Price Adjustment with Menu Cost for Multi-product Firms," *Manuscript*, 2010.

_ , _ , and Luigi Paciello, "Optimal price setting with observation and menu costs," *The Quarterly Journal of Economics*, 2011, *126* (4), 1909–1960.

Bachmann, Rüdiger and Giuseppe Moscarini, "Business cycles and endogenous uncertainty," in "2011 Meeting Papers," Vol. 36 Society for Economic Dynamics 2011.

_ , Benjamin Born, Steffen Elstner, and Christian Grimme, "Time-varying business volatility and the price setting of firms," *Journal of Monetary Economics*, 2019, *101*, 82–99.

Baley, Isaac and Julio A Blanco, "Menu costs, uncertainty cycles, and the propagation of nominal shocks," *American Economic Journal: Macroeconomics*, 2018, *forthcoming.*

Bloom, Nicholas, "The impact of uncertainty shocks," *econometrica*, 2009, *77* (3), 623–685.

_ , Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J Terry, "Really uncertain business cycles," *Econometrica*, 2018, *86* (3), 1031–1065.

Caplin, Andrew and John Leahy, "Aggregation and optimization with state-dependent pricing," *Econometrica*, 1997, pp. 601–625.

_ , Mark Dean, and John Leahy, "Rational inattention, optimal consideration sets, and stochastic choice," *The Review of Economic Studies*, 2019, *86* (3), 1061–1094.

Coibion, Olivier and Yuriy Gorodnichenko, "Information rigidity and the expectations formation process: A simple framework and new facts," *The American Economic Review*, 2015, *105* (8), 2644–2678.

_ , _ , and Saten Kumar, "How do firms form their expectations? new survey evidence," *American Economic Review*, 2018, *108* (9), 2671–2713.

Costain, James and Anton Nakov, "Precautionary price stickiness," *Journal of Economic Dynamics and Control*, 2015, *58*, 218–234.

_ and _ , "Logit price dynamics," *Journal of Money, Credit and Banking*, 2019, *51* (1), 43–78.

Drenik, Andres and Diego Perez, "Price setting under uncertainty about inflation," *Working Paper*, 2018.

Farmer, Leland E, "The Discretization Filter: A Simple Way to Estimate Nonlinear State Space Models," *SSRN Working Papers*, 2016.

_ **and Alexis Akira Toda**, "Discretizing nonlinear, non-Gaussian Markov processes with exact conditional moments," *Quantitative Economics*, 2017, *8* (2), 651–683.

Gorodnichenko, Yuriy, "Endogenous information, menu costs and inflation persistence," Technical Report, National Bureau of Economic Research 2008.

Hellwig, Christian and Laura Veldkamp, "Knowing what others know: Coordination motives in information acquisition," *The Review of Economic Studies*, 2009, *76* (1), 223–251.

Jung, Junehyuk, Jeong Ho Kim, Filip Matějka, and Christopher A Sims, "Discrete actions in information-constrained decision problems," *The Review of Economic Studies*, 2019, *86* (6), 2643–2667.

Klenow, Peter J and Oleksiy Kryvtsov, "State-dependent or time-dependent pricing: Does it matter for recent US inflation?," *The Quarterly Journal of Economics*, 2008, *123* (3), 863–904.

Klepacz, Matthew, "Price setting and volatility: Evidence from oil price volatility shocks," *Working Paper*, 2017.

Kumar, Saten, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko, "Inflation targeting does not anchor inflation expectations: Evidence from firms in New Zealand," Technical Report, National Bureau of Economic Research 2015.

Maćkowiak, Bartosz and Mirko Wiederholt, "Optimal sticky prices under rational inattention," *The American Economic Review*, 2009, *99* (3), 769–803.

_ **, Filip Matějka, and Mirko Wiederholt**, "Dynamic rational inattention: Analytical results," *Journal of Economic Theory*, 2018, *176*, 650–692.

Matějka, Filip, "Rationally inattentive seller: Sales and discrete pricing," *The Review of Economic Studies*, 2015, p. rdv049.

Matejka, Filip and Alisdair McKay, "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *The American Economic Review*, 2014, *105* (1), 272–298.

Miao, Jianjun and Hao Xing, "Dynamic Rationally Inattentive Discrete Choice: A Posterior-Based Approach," Technical Report, Boston University-Department of Economics 2019.

Midrigan, Virgiliu, "Menu costs, multiproduct firms, and aggregate fluctuations," *Econometrica*, 2011, *79* (4), 1139–1180.

Moscarini, Giuseppe, "Limited information capacity as a source of inertia," *Journal of Economic Dynamics and control*, 2004, *28* (10), 2003–2035.

Nakamura, Emi and Jón Steinsson, "Five facts about prices: A reevaluation of menu cost models," *The Quarterly Journal of Economics*, 2008, *123* (4), 1415–1464.

Nieuwerburgh, Stijn Van and Laura Veldkamp, "Learning asymmetries in real business cycles," *Journal of monetary Economics*, 2006, *53* (4), 753–772.

Paciello, Luigi and Mirko Wiederholt, "Exogenous Information, Endogenous Information, and Optimal Monetary Policy," *Review of Economic Studies*, 2013, *81* (1), 356–388.

Shannon, CE, "A mathematical theory of communication," *The Bell System Technical Journal*, 1948, *27* (3), 379–423.

Sims, Christopher A, "Implications of rational inattention," *Journal of monetary Economics*, 2003, *50* (3), 665–690.

Steiner, Jakub, Colin Stewart, and Filip Matějka, "Rational Inattention Dynamics: Inertia and Delay in Decision-Making," *Econometrica*, 2017, *85* (2), 521–553.

Stevens, Luminita, "Coarse pricing policies," *Review of Economic Studies*, 2019, *forthcoming.*

Tutino, Antonella, "Rationally inattentive consumption choices," *Review of Economic Dynamics*, 2013, *16* (3), 421–439.

Vavra, Joseph, "Inflation dynamics and time-varying volatility: New evidence and an ss interpretation," *The Quarterly Journal of Economics*, 2013, *129* (1), 215–258.

Wiederholt, Mirko, "Rational inattention," *The New Palgrave Dictionary of Economics (Online Edition ed.)*, 2010.

Woodford, Michael, "Imperfect common knowledge and the effects of monetary policy," *In Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps, ed. Philippe Aghion, Roman Frydman, Joseph Stiglitz, and Michael Woodford*, 2003.

_ , "Information-constrained state-dependent pricing," *Journal of Monetary Economics*, 2009, *56*, S100–S124.

7 Appendix

7.1 Appendix A: profit function approximation

The derivation follows closely Alvarez and Lippi (2010). All firms share the same profit function $\Pi(P_t, Y_t, C_t) = Y_t P_t^{-\eta}(P_t - C_t)$. Where $\eta > 1$ represents the constant price elasticity, Y_t is the intercept of the demand (i.e. it's a demand shifter) and C_t is the marginal cost at time t. I assume that Y_t and C_t are perfectly correlated, i.e. when costs are high demand is also high. In order to approximate the objective function as (1), I compute a second order approximation of $\Pi(P_t, Y_t, C_t)$ around its frictionless price. In the RI context, the frictionless price is the optimal price under full information P_t^*.

The second order approximation of $\Pi(P_t, Y_t, C_t)$

$$\Pi(P_t, Y_t, C_t) \approx \Pi(P_t^*, Y_t, C_t) + \left.\frac{\partial \Pi(P_t, Y_t, C_t)}{\partial P_t}\right|_{P_t = P_t^*}(P_t - P_t^*) + \frac{1}{2}\left.\frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2}\right|_{P_t = P_t^*}(P_t - P_t^*)^2$$

Which can be written:

$$\begin{aligned}\frac{\Pi(P_t, Y_t, C_t)}{\Pi(P_t^*, Y_t, C_t)} = 1 + \frac{1}{\Pi(P_t^*, Y_t, C_t)}\left.\frac{\partial \Pi(P_t, Y_t, C_t)}{\partial P_t}\right|_{P_t = P_t^*} P_t^* \frac{(P_t - P_t^*)}{P_t^*} \\ + \frac{1}{2}\frac{1}{\Pi(P_t^*, Y_t, C_t)}\left.\frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2}\right|_{P_t = P_t^*}(P_t^*)^2 \left(\frac{P_t - P_t^*}{P_t^*}\right)^2\end{aligned}$$

Taking the first and second order conditions:

$$\begin{aligned}\frac{\partial \Pi(P_t, Y_t, C_t)}{\partial P_t} &= Y_t P_t^{-\eta}\left[-\eta\left(\frac{P_t - C_t}{P_t}\right) + 1\right] \\ \frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2} &= -Y_t P_t^{-\eta-1}\eta\left[-\eta\left(\frac{P_t - C_t}{P_t}\right) + 1\right] - Y_t \eta P_t^{-\eta-2} C_t\end{aligned}$$

From the first order conditions, the optimal price is simply a constant mark-up over marginal cost $P_t = \frac{\eta}{\eta-1}C_t$. Evaluating the first and second order conditions at the optimal price:

$$\begin{aligned}\left.\frac{\partial \Pi(P_t, Y_t, C_t)}{\partial P_t}\right|_{P_t^*} &= 0 \\ \left.\frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2}\right|_{P_t^*} &= -\eta Y_t C_t \left(\frac{1}{P_t^*}\right)^2 \left(\frac{\eta}{\eta - 1}C_t\right)^{-\eta}\end{aligned}$$

The maximized value of the profits:

$$\Pi(P_t^*, Y_t, C_t) = Y_t \left(\frac{\eta}{\eta - 1}\right)^{-\eta} C_t^{1-\eta} \left(\frac{1}{\eta - 1}\right)$$

Therefore, the term:

$$\frac{1}{2}\frac{1}{\Pi(P_t^*, Y_t, C_t)}\frac{\partial^2\Pi(P_t, Y_t, C_t)}{\partial P_t^2}\bigg|_{P_t}(P_t^*)^2 = \frac{-\eta Y_t C_t \left(\frac{\eta}{\eta-1}C_t\right)^{-\eta}}{Y_t \left(\frac{\eta}{\eta-1}\right)^{-\eta} C_t^{1-\eta} \left(\frac{1}{\eta-1}\right)} = -\eta(\eta - 1)$$

Finally, the second order approximation:

$$\frac{\Pi(P_t, Y_t, C_t) - \Pi(P_t^*, Y_t, C_t)}{\Pi(P_t^*, Y_t, C_t)} = -\frac{1}{2}\eta(\eta - 1)\left(\frac{P_t - P_t^*}{P_t^*}\right)^2 + o\left(\frac{P_t - P_t^*}{P_t^*}\right)$$

Where I can finally define $\gamma \equiv -\frac{1}{2}\eta(\eta-1)$, $\widehat{\Pi}(p_{it}, \widehat{p}_{it}) = log(\Pi(P_t, Y_t, C_t)) - log(\Pi(P_t^*, Y_t, C_t))$, $p_t = log(P_t)$ and $\widehat{p}_{it} = log(P_t^*)$ as stated in equation (1).

7.2 Appendix B: equivalence of mutual information

Information entropy is a measure about the uncertainty of a random a variable. Consider a random variable X with finite support Ω_x, which is distributed according to $f \in \Delta(\Omega_x)$. The entropy of X, is defined by:

$$\mathcal{H}(X) = -\sum_{x\in\Omega_s} f(x) log f(x)$$

With the convention that $0 log 0 = 0$. In RI, the acquired amount of information is measured by entropy reduction. Given the signal s_t, entropy reduction is measured by mutual information, which in the context of this dynamic model is:

$$\mathcal{I}(\widehat{p}_{it}, s_{it}|s_i^{t-1}) = \mathcal{H}(\widehat{p}_{it}|s_i^{t-1}) - E_{s_{it}}[\mathcal{H}(\widehat{p}_{it}|s_{it})|s_i^{t-1}]$$

Given the entropy, the target-price $\widehat{p}_{it} = \sigma_t \epsilon_{it} \in \Omega_{\widehat{p}}$, and the definition for mutual information we can prove:

$$\begin{aligned}
\mathcal{I}(\widehat{p}_{it}, s_{it}|s_i^{t-1}) &= \mathcal{H}(\widehat{p}_{it}|s_i^{t-1}) - E_{s_{it}}[\mathcal{H}(\widehat{p}_{it}|s_{it})|s_i^{t-1}] \\
&= \sum_{s_{it}} f(s_{it}|s_i^{t-1}) \left[\sum_{\sigma}\sum_{\epsilon} f(\widehat{p}_{it}|s_{it}, s_i^{t-1}) log(f(\widehat{p}_{it}|s_{it}, s_i^{t-1})) \right] \\
&- \sum_{\sigma_t}\sum_{\epsilon_{it}} g(\widehat{p}_t|s_i^{t-1}) log(g(\widehat{p}_t|s_i^{t-1})) \\
&= \sum_{s_{it}}\sum_{\sigma_t}\sum_{\epsilon_{it}} f(s_{it}, \widehat{p}_{it}|s_i^{t-1}) log(f(\widehat{p}_{it}|s_{it}, s_i^{t-1})) \\
&- \sum_{\sigma_t}\sum_{\epsilon_{it}} \left[\sum_{s_{it}} f(s_{it}, \widehat{p}_{it}|s_i^{t-1}) \right] log(g(\widehat{p}_{it}|s_i^{t-1})) \\
&= \sum_{s_{it}}\sum_{\sigma_t}\sum_{\epsilon_{it}} f(s_{it}, \widehat{p}_{it}|s_i^{t-1}) log\left(\frac{f(\widehat{p}_{it}|s_{it}, s_i^{t-1})}{g(\widehat{p}_{it}|s_i^{t-1})} \right) \\
&= \sum_{s_{it}}\sum_{\sigma_t}\sum_{\epsilon_{it}} f(s_{it}, \widehat{p}_{it}|s_i^{t-1}) log\left(\frac{f(s_{it}, \widehat{p}_{it}|s_i^{t-1})}{g(\widehat{p}_{it}|s_i^{t-1}) f(s_{it}|s_i^{t-1})} \right)
\end{aligned}$$

Using the notation $\sum_{x_t} = \sum_{x_t \in \Omega_x}$.

From the second to the third line of the equivalence we rely on the fact that the prior distribution (marginal) is characterized as the sum of the joint probability distribution $f(s_{it}, \widehat{p}_{it}|s_i^{t-1})$ across all potential signals. The final expression is then what is shown in equation (3).

7.3 Appendix C: solution of the dynamic RI problem

In this section, I show how to derive the solution for the dynamic RI problem formally introduced in Section 2.4. Given prior beliefs $g_{it}(\widehat{p}_{it}|p_{it-1})$, firms choose the conditional probability distribution of prices $f_{it}(p_{it}|\widehat{p}_{it})$ (equivalent of choosing $f(p_{it}, \widehat{p}_{it})$) in each point of the simplex $\Omega_p \times \Omega_\sigma \times \Omega_\epsilon$. To simplify notation, I will omit the lagged price conditioning and focus on a representative firm $\lambda_i = \lambda$.

Since the prior belief about the volatility distribution $m_t(\sigma_L)$ is the state variable of the problem, we can write the Bellman equation:

$$V(m_t(\sigma_L)) = \max_{f_t(p_t|\widehat{p}_t)} \sum_\sigma \sum_\epsilon \sum_p [\widehat{\Pi}(p_t, \widehat{p}_t) + \beta V(m_{t+1}(\sigma_L))] f_t(p_t|\widehat{p}_t) g_t(\widehat{p}_t) \ - \ \lambda \mathcal{I}(\widehat{p}_t, p_t)$$

Where:

$$\mathcal{I}(\widehat{p}_t, p_t) = f_t(p_t, \widehat{p}_{it}) log \left(\frac{f_t(p_t, \widehat{p}_t)}{g_t(\widehat{p}_t) f_t(p_t)} \right) = f_t(p_t|\widehat{p}_t) g_t(\widehat{p}_t)[log(f_t(p_t|\widehat{p}_t)) - log(f_t(p_t))]$$

The function is also maximized subject to the constraint on the prior (7). The first order condition of $V(m_t(\sigma_L))$ with respect to $f_t(p_t|\widehat{p}_{it})$:

$$\begin{aligned} g_t(\widehat{p}_t) \left[\widehat{\Pi}(p_t, \widehat{p}_{it}) + \beta V(m_{t+1}(\sigma_L)) + \beta \left[\frac{\partial V(m_{t+1}(\sigma_L))}{\partial m_{t+1}(\sigma_L)} \times \frac{\partial m_{t+1}(\sigma_L)}{\partial f_t(p_t|\widehat{p}_t)} \right] \right] \\ - \lambda g_t(\widehat{p}_t)[log(f_t(p_t|\widehat{p}_t)) + 1 - log(f_t(p_t)) - 1] - g_t(\widehat{p}_t)\mu(\widehat{p}_t) = 0 \end{aligned} \tag{18}$$

The last term on the left hand side of equation (18), $\mu(\widehat{p}_{it})$, corresponds to the Lagrange multiplier of the constraint attached to the prior, equation (7).

Embedded in equation (18) is the effect of the current information strategy on posterior beliefs, $\frac{\partial m_{t+1}(\sigma_L)}{\partial f_t(p_t|\widehat{p}_t)}$. As discussed, posterior beliefs will later become the prior for $t+1$, $g_{t+1} = m_{t+1}(\sigma)h(\epsilon)$. The known i.i.d. structure of the idiosyncratic shocks ϵ_t implies that the chosen information strategy is not going to affect beliefs about this marginal distribution. Moreover, as stressed by SSM (2017), we can treat the effects of current information on future beliefs about the persistent state σ_t as fixed. The authors shows that a dynamic RI problem such as the one presented in this paper, is equivalent to a control problem without uncertainty about persistent states.[22] Therefore $\frac{\partial m_{t+1}(\sigma_L)}{\partial f_t(p_t|\widehat{p}_t)} = 0$ and given $g_t(\widehat{p}_t) \geq 0$ and $\lambda > 0$, equation (18) becomes:

$$\begin{aligned} \frac{\Pi(p_t, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L)) - \mu(\widehat{p}_t)}{\lambda} &= \ log\left(\frac{f(p_t|\widehat{p}_t)}{f_t(p)}\right) \\ exp\left(\frac{\Pi(p_t, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda}\right) exp\left(\frac{-\mu(\widehat{p}_t)}{\lambda}\right) &= \frac{f(p_t|\widehat{p}_t)}{f_t(p)} \\ \Rightarrow f(p_t|\widehat{p}_t) = \ exp\left(\frac{\Pi(p_t, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda}\right) & f_t(p_t)\phi(\widehat{p}_t) \end{aligned}$$

Where:

$$\phi(\widehat{p}_t) \equiv exp\left(\frac{-\mu(\widehat{p}_t)}{\lambda}\right) \tag{19}$$

Finally, due to the restriction on the prior:

$$\begin{aligned} g_t(\widehat{p}_t) &= \sum_{p_t^{'}} f_t(p_t^{'}|\widehat{p}_t)g(\widehat{p}_t) \\ &= \sum_{p_t^{'}} exp\left(\frac{\Pi(p_t^{'}, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda}\right) f_t(p_t^{'})\phi(\widehat{p}_t)g(\widehat{p}_t) \\ \Rightarrow \phi(\widehat{p}_t) &= \frac{1}{\sum_{p_t^{'}} \ exp\left(\frac{\Pi(p_t^{'}, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda}\right) f_t(p_t^{'})} \end{aligned}$$

Combining this expression with (19), and adding the conditioning on lagged prices, we get the expression for the optimal posterior distribution of prices given the unobserved target, (11):

$$f_t(p_t|\widehat{p}_t, p_{t-1}) = \frac{exp\left[\left(\Pi(p_t, \widehat{p}_t) + \beta V(m_{t+1}(\sigma_L|p_t))\right)/\lambda\right] f_t(p_t|p_{t-1})}{\sum_{p_t^{'}} \ exp\left[\left[\Pi(p_t^{'}, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L|p_t))\right)/\lambda\right] f_t(p_t^{'}|p_{t-1})}$$

[22]The intuition behind the result is the following. In the control problem, while firms have full information about the current and past history of shocks, they face a trade-off of optimizing their flow utility $\widehat{\Pi}(p_t, \widehat{p}_t)$ against a control cost given by: $E_{f(p_t|\widehat{p}_{it})}[log(f(p_t|\widehat{p}_t)) - log(q(p_t|\widehat{p}_t)|z^t]$. The variable z^t stands for the entire history of past shocks and prices. The cost is determined by the deviation of the final action with respect to some default action $q(p_t|\widehat{p}_{it})$. By relying on properties of the entropy, the paper shows an equivalence between a control and a dynamic Rational Inattention problem. Thus the inattention problem is solved by initially solving the control problem with observable states, characterizing the optimal conditional probability for each default rule $f(p_t|\widehat{p}_t)$, and then choosing q. As states are observable in the control problem, the solution ignores the effects of information acquisition on future beliefs (i.e., treat them as fixed) when solving the dynamic RI problem.

The expression for the value function is then simply given by plugging this expression into (5):

$$
\begin{aligned}
V(m_t(\sigma_L)) &= \lambda \sum_{\sigma_t} \sum_{\epsilon_t} \sum_{p_t} f(p_t, \widehat{p}_t) \log \left(\sum_{p} exp \left(\frac{\Pi(p_t, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda} \right) f(p_t|p_{t-1}) \right) \\
&= \lambda E \left[\log \left(\sum_{p_t} \ exp \left(\frac{\Pi(p_t, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda} \right) f(p_t|p_{t-1}) \right) \right]
\end{aligned}
$$

7.4 Appendix D: Dynamic RI Algorithm

The algorithm to solve the dynamic RI problem is as follows:

1. Fix a value for the idiosyncratic information acquisition cost, e.g. λ_1.

2. Given λ_1 and the belief simplex, compute prior beliefs $g(\widehat{p}_{it}) = m(\sigma_t)h(\epsilon_{it})$.

3. With $g(\widehat{p}_{it})$, the model is solved by Value Function Iteration.

   3.1. Starting with a guess for the vector $V(m_{t+1}(\sigma_L))$, we first solve the static RI problem. The algorithm computes $f(p_{it}, \widehat{p}_{it}|p_{it-1}) \in \Delta(\Omega_p \times \Omega_\sigma \times \Omega_\epsilon)$ which is the solution for the system of nonlinear equations (7), (11) and $f(p_{it}|p_{it-1}) = \sum_\sigma \sum_\epsilon f(p_{it}, \widehat{p}_{it}|p_{it-1})$.

   3.2. Given $f(p_{it}, \widehat{p}_{it}|p_{it-1})$, the prior $g_{it}(\widehat{p}_{it})$ and using Bayes Law, we can compute the conditional probability $f(\sigma|p_{it}, p_{it-1}) = \sum_\epsilon f(\sigma, \epsilon|p_{it}, p_{it-1})$ for each $p_{it} \in \Omega_p$. Through (10), posterior beliefs become the prior beliefs for the next period. With this we update $V(m_{t+1}(\sigma_L))$.

   3.3. Relying on the definition for $V(m_{it}(\sigma_L|p_t))$ in (12), the algorithm iterates the value function until convergence when, within each iteration, it re-estimates $f(p_{it}, \widehat{p}_{it}|p_{it-1})$.

4. Repeat point 3 for all possible values in $\Delta(\Omega_\sigma)$, i.e. setting different priors $g(\widehat{p}_{it})$.

5. Repeat 2, 3, and 4 for all possible values for λ_i.

The setting of the model and the decision of the shape of the joint probability distribution resembles a filtering problem. The numerical discrepancies between filtering with discrete variables relative to continuous outcomes are not significant and depend on the nature of the approximation, Farmer (2016) and Farmer and Toda (2017).

7.5 Appendix E: Sensitivity analysis

In this section, we provide a further description of the identification strategy of our baseline model. As discussed we targeted four moments, $E(|\Delta p|)$, $E(Dispersion|\sigma_L)$, $E(Dispersion|\sigma_H)$ and β_{IR}, using for parameters $\theta = \{\sigma_L, \phi, \overline{\lambda}, \sigma_\lambda\}$.

To provide a sensitivity analysis of which data moments are more informative of which parameters we perform the following exercise: starting from the first parameters in θ, i.e., σ_L, we reduce its magnitude by 0.1% while keeping *all* the remaining parameters fixed at their calibrated values. With this new parametrization we solve the model again, and report the four targeted moments. We repeat this exercise for different values of σ_L, where its original magnitude is reduced or increased by $0.x\%$ where $x = 1, \ldots, 5$. We repeat the procedure for each parameter in θ one at a time, keeping all of the remaining parameters fixed in their original calibrated values. The results are shown in Figure 6.

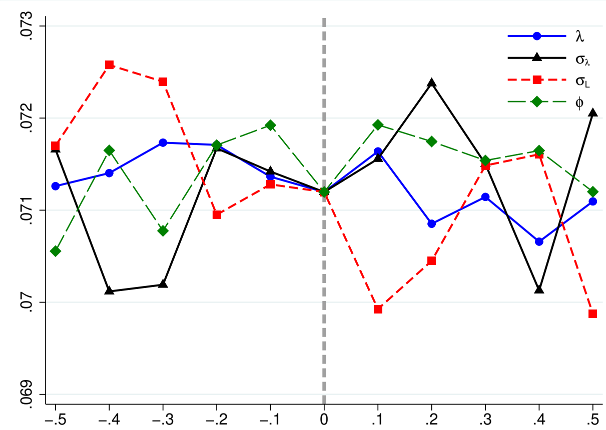
Although we allow for some marginal perturbations of the parameters, in some cases the response of the targeted moments is sizable. We conjecture that this is because of the interplay between the effects of the new parameters on the learning and pricing strategies combined with the discretization assumed for the simplex of each variable.

The average magnitude of price revisions $E(|\Delta p|)$ (top left panel of Figure 6) seems very sensible to volatility changes in the more persistent (and therefore the most likely) aggregate state. Changes in σ_L, represented by the connected line with squares, affects the intensive margin of prices significantly. This moment is also very sensible to the dispersion of information costs, shown by the black line with triangles. In line with was described in Section 3.3, heterogeneous costs leads to heterogeneous pricing strategies creating a direct mapping between the array of values for λ_i and the magnitude of price adjustments. The degree by which $E(|\Delta p|)$ is affected by σ_λ is simply a direct implication of this result.

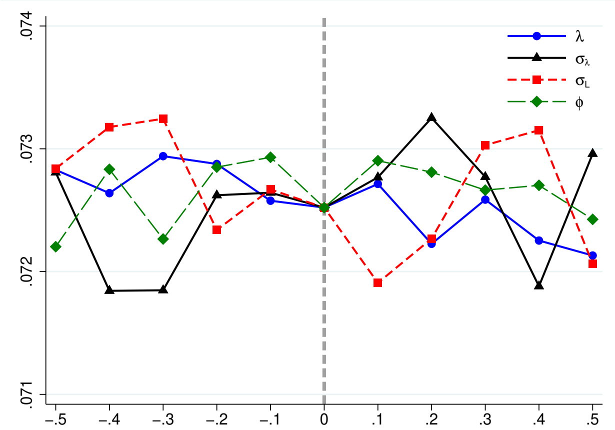
As the dispersion of information costs σ_λ disciplines the relative magnitude of price adjustments between firms, the two dispersion moments (top right and bottom left figures) are also highly responsive to this parameter. As expected, the price dispersion in the low volatility state is also affected by σ_L. Intuitively, the relevance of this last parameter is more muted when we look at the reaction of price dispersion in the high volatility state. In this latter case, and by construction, the role of ϕ in identifying this moment is more relevant relative to σ_L. However, for these two moments the implications are not so clear, suggesting that they are jointly identified by the set of parameters.

Finally, the overall degree of information rigidity β_{IR} in the model seems very sensible to average value of information costs $\overline{\lambda}$. As discussed in the main text, this parameter is identified by regressing the forecast error on the forecast revision over the cross-section of firms. As shown by CG (2015) there is a direct mapping between this parameter and the degree of information rigidity faced by agents. As $\overline{\lambda}$ reflects the average magnitude of the information rigidity across firms it is not surprising that this moment is very responsive to this parameter.

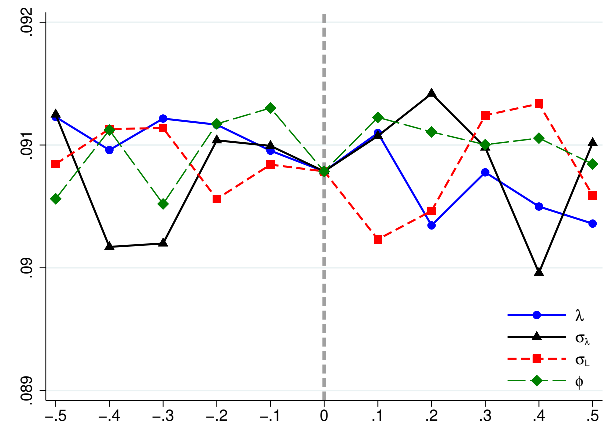
Figure 6: Sensitivity of moments to parameters



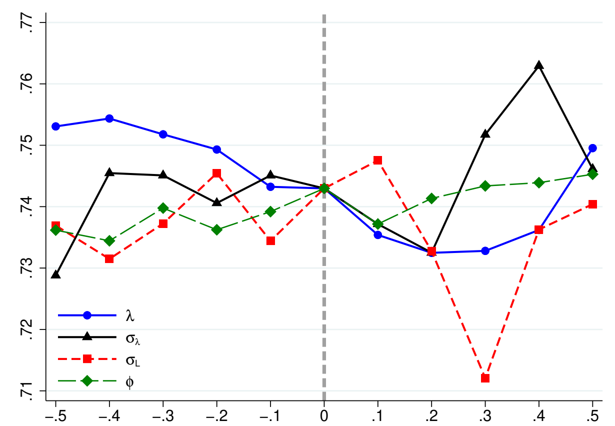
(a) $E(|\Delta p|)$



(b) $E(Dispersion|\sigma_L)$



(c) $E(Dispersion|\sigma_H)$



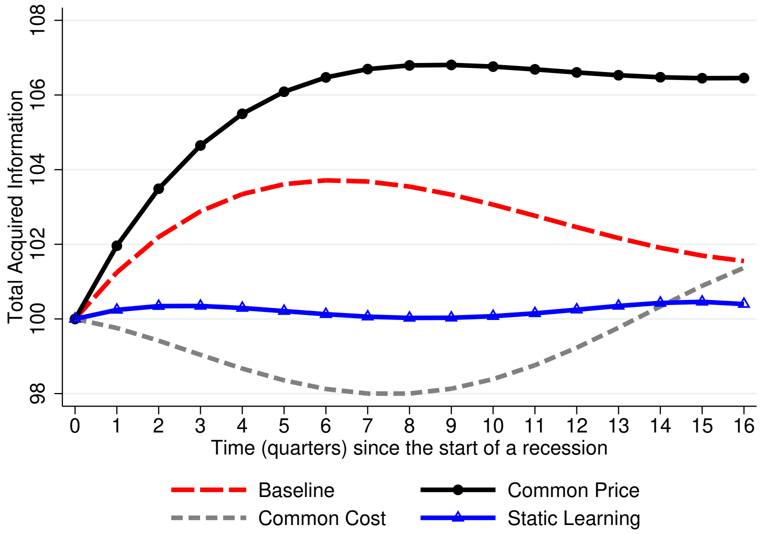
(d) β_{IR}

Notes: The figure presents the sensitivity of the four targeted moments to changes in one of the parameters in θ, while keeping the rest constant. The procedure is repeated for all four parameters where within each iteration the dynamic learning model is solved and the relevant moments are saved.

7.6 Appendix F: Information acquisition - Alternative Models

In this section we study the different implications for dynamic learning over the business cycle for all alternative specifications presented in Section 5. Figure 7 shows the simulated evolution for the baseline model along with the static, homogeneous price, and common cost models. As in the original figure 1, we show the response of total acquired information after the economy enters into the high volatility state at quarter $t = 0$. Since the focus in on the underlying dynamics, we normalize the initial response to 100.

Figure 7: Information acquisition over the business cycle - Alternative Models



Notes: The figure represents the impulse response function of the different simulated models after the economy enters into the less predictable state at time $t = 0$. The red dashed line represents the baseline model, the blue with triangles is the static learning model, the dotted black lines is the common price version and finally the dashed gray line is the homogeneous cost specification. The initial responses are normalized to 100.

7.6.1 Static Learning

Despite the muted predictions for β_{IR}, we can still assess if the static learning model is consistent with state-dependent attention and its dynamic evolution. Based on equation (15) it is straightforward to notice that total attention will increase immediately after the economy enters into the less predictable state at $t = 0$. As expected, the figure shows that κ_t will stays almost constant throughout the 16 simulated quarters. This result is completely in line with the average duration of the high volatility state given by the calibrated transition probabilities.[23]

[23] Actually, the two transitions probabilities $\tau_{LH} = 0.00882$ and $\tau_{HL} = 0.0196$ imply that the average duration of the high volatility state is 17 quarters.

While the initial response is normalize to 100, the magnitude of $\overline{\kappa}$ in this case is significantly higher than in the other cases. In particular, $\overline{\kappa}$ is around 8 times higher compared to the baseline model. In the static version, price-setters does not waste any of their attention in noticing the actual state of the economy. Hence, the learning problem is simpler as firms only collect information to track the outcome of the target-price, leading them to rationally choose to collect more information.

The implied dynamics for κ in this case are completely at odds with the dynamic patters of both the data and the baseline model. This, along with the discussion in Section 5.5.1, reinforces the impossibility of a more stylized version of the model to consistently replicate the features of the data.

7.6.2 Common Cost

Given the same cost of information λ, the learning and pricing strategies will be common across firms. Hence, the rate by which firms will notice any new aggregate state will also be similar across them. Consistent with the results discussed in Section 3.3, a firm with an average cost will focus almost all of its attention on a subset of prices which are closer to the mean. This would prevent them from noticing any extreme realization of the target-price that could suggest that the economy is actually in the more volatile state. This intuition is reinforced by the results in figure 7. The fact that $\overline{\kappa}$ slightly decreases until the 7-8th quarter after the recession starts suggest that firms confound the new state with the more predictable state, leading them to marginally collect less information. It is only after several quarters of being exposed to more extreme realization of $\widehat{p}_{it}$ that they start revising their beliefs in favor of the high volatility state. The rise in $\overline{\kappa}$ is therefore delayed due to their misperceived beliefs.

The simulated response of total information is not only inconsistent with the data, but also reinforces the relevance of allowing for idiosyncratic differences at the firm level as a relevant mechanism to capture the time-varying implication of attention.

7.6.3 Common Target Price

Among the alternative models, the version with a common optimal price successfully resembles the dynamic features of the data (black dotted line in figure 7). Although this version of the model is not fully consistent with some key targeted price facts, this result highlights the relevance of having both dynamic information and firm heterogeneity to replicate state-dependent attention.

7.7 Appendix G: Robustness to first and second-moment shocks

This section provides further detailed results for the extended version of the model with both first and second-moments shocks. As discussed in Section 5.6, the consequences of adding a first-moment shock to our model are not obvious, particularly within a setup where information is endogenous and fully flexible. Since the entire learning strategy will change with the two shocks, it is hard to anticipate the changes in acquired information, which complicates the comparison with our baseline scenario. Hence, we solve the two shocks version using the original parameters in Table 2 and setting $\mu_H = 0.995\mu_L$. Table 4 shows the pricing and the information moments.

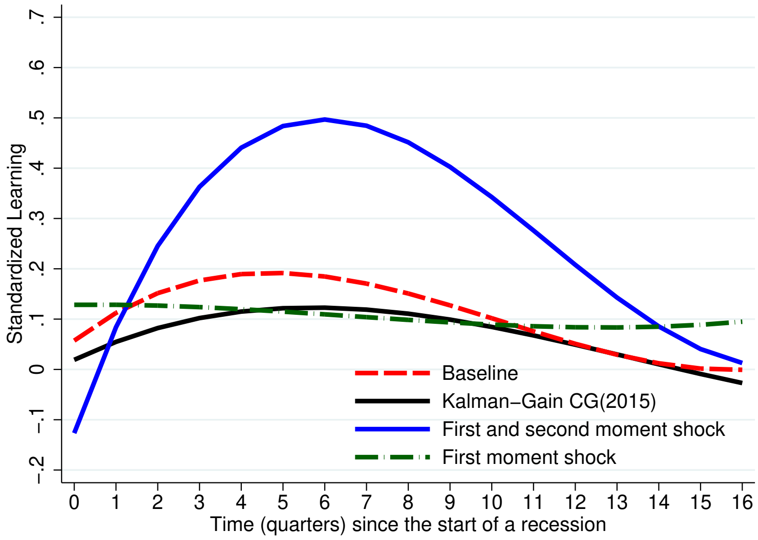
Table 4: Matched Moments and Alternative Specifications

Targeted moments	Data	Baseline	First & Second Moment
$E(\|\Delta p\|)$	0.077	0.071	0.061
$E(Dispersion\|\sigma_L)$	0.073	0.072	0.068
$E(Dispersion\|\sigma_H)$	0.090	0.090	0.086
β_{IR}	0.674	0.724	0.773
Non-Targeted moments			
Frequency	0.150	0.564	0.636
$Kurtosis(\|\Delta p\|)$	6.403	4.049	4.127
Fraction small	0.330	0.175	0.209
$Corr(Dis, Freq)$	0.506	0.630	0.516

Although the moments are not directly targeted, the distance between the extended model and the data is not far. This is reassuring as it implies that the results are robust to more complex versions of the dynamic learning model, accounting for additional business cycle features. The model can still replicate (and even be closer to match) one of the key non-targeted moments, such as the correlation between dispersion and frequency of price adjustments. Turning to the learning responses, Figure 8 shows the standardized response of acquired information over the cycle. From the Figure, we notice that, the presence of the two shocks brings an amplification in the learning reaction of firms relative to the scenario with just a second-moment shock. Intuitively, the different mean makes the learning problem easier for firms as now the perceived average of signals changes over time. This boosts learning right after the economy enters into a recession. Besides the different learning rates, the dynamic correlation between the extended model and the data is 0.83, a marginal improvement relative to the 0.79 correlation of the baseline data.[24]

[24] We conjecture that by calibrating the two-shock version of the model to match the targeted moments, both the empirical and the simulated responses would be closer to each other. This is because one of the targeted moments is the average information rigidity parameter β_{IR}. However, this section intended to address whether a more sophisticated model robust to business cycle features can be consistent with the data. We leave the challenge of calibrating such a model for future work.

Figure 8: IRF Response



7.7.1 First-moment shock only

Finally, we propose a different version that allows for a drop in the mean while keeping the volatility constant. In particular, we assume that $\widehat{p}_{it} = \mu_t + \sigma_t \epsilon_{it}$ where $\mu_H < \mu_L$ and $\sigma_t = \sigma_L$. The response of this specification to a recession is shown by the green dotted line in Figure 8. Allowing for just a first-moment shock is not enough to generate any meaningful reaction in the learning rate. Although the states of the economy are still persistent, the only evidence suggesting a change of state is given by a shift in the average of acquired signals. Since the volatility of the target price remains constant, firms' learning problem is possibly more straightforward than the time-varying volatility setting. The level of total learning stays relatively constant over time. In particular, it randomly revolves around the initial attention level of 0.1 approximately.

As shown in Section 5.1 and consistent with Wiederholt (2010), in RI models, more attention is endogenously devoted to more volatile processes. Within a price-setting model, Maćkowiak and Wiederholt (2009) shows that the different attitudes towards volatility can explain why prices react more quickly to idiosyncratic relative to aggregate shocks. As idiosyncratic conditions are more volatile relative to aggregate ones, price-setters respond by focusing more attention on the former than the latter shock. The results with only a mean shock point exactly in this direction. While the results are robust to having both a mean and a variance shock, it is the presence of a rise in volatility that endogenously boosts learning over the business cycle, resembling the data.